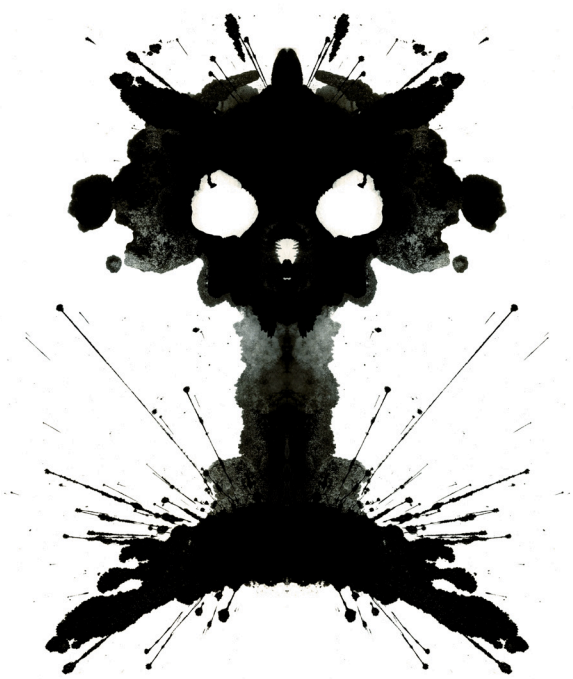
KTH ROYAL INSTITUTE
OF TECHNOLOGY

Doctoral Thesis in Philosophy

Deterrence Games for the 21st Century

Representation, Theory and Evidence

KARL SÖRENSON



Stockholm, Sweden 2022

Deterrence Games for the 21st Century

Representation, Theory and Evidence

KARL SÖRENSON

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology, is submitted for public defence for the Degree of Doctor of Philosophy on Tuesday the 17th of May 2022, at 15:00 (GMT+1) online/zoom, Stockholm.

Doctoral Thesis in Philosophy
KTH Royal Institute of Technology
Stockholm, Sweden 2022

ISBN 978-91-8040-101-2
TRITA-ABE-DLT-2147

Printed by: Universitetsservice US-AB, Sweden 2022

Contents

Abstract

KARL SÖRENSON, *Deterrence Games for the 21st Century: Representation, Theory and Evidence*, Doctoral thesis in Philosophy. *Theses in Philosophy from the Royal Institute of Technology* 68. Stockholm, 2021, 140 pp. ISBN 978-91-8040-101-2. TRITA-ABE-DLT-2147.

Deterrence is the backbone of military strategy. Dissuading an opponent from taking a specific action by threat of violence is the definition of deterrence. From the outset of the Cold War there has been a strong link between the study of deterrence and game theoretic analysis. There are compelling epistemic reasons for studying deterrence as a game. By doing so, the strategic interaction between actors is placed at the centre of the analysis, mapping the possible outcomes and revealing the strategies available to the actors. Discussions about various models' appropriateness and model comparison therefore play a central role in deterrence research; from underlying assumptions and deterrence representation to theory and evidence. This dissertation treats aspects of all of these topics. Article I, "Prospects of Deterrence – Deterrence Theory – Representation and Evidence", analyses the relationship between model and theory and what happens to a deterrence theory when the rationality assumption is switched to a prospect theoretical utility function. Article II, "A Misfit Model – Bounded Rationality and Deterrence Representation", defends and remodels Schelling's idea of irrational threats for effective deterrence. Article III, "Comparable Deterrence – Target, Criteria and Purpose", treats the issue of how one can compare game theoretic models with one another and proposes a meta-model for how this can be done. Article IV, "Deterring the Dauntless – Appraising the Effects of Naval Deterrence against the Somali Piracy", estimates whether and to what extent Somali piracy was deterred by the naval intervention.

Thesis composition

This thesis consists of an introduction and the following four research articles:

I. Prospect of Deterrence: Deterrence Theory, Representation and Evidence.

II. A Misfit Model: Bounded Rationality and Deterrence Representation.

III. Comparable Deterrence: Target, Criteria and Purpose.

IV. Deterring the Dauntless: Appraising the Effects of Naval Deterrence Against the Somali Piracy.

In the introduction, the research articles will be referenced with their Roman numerals as per above.

Author's Note

A part of my family's lineage traces its roots to the Stockholm archipelago, the vast chain of islands surrounding the Swedish capital, and particularly the isle of Sandön. As child, I spent many summers on and around that very island. I remember being fascinated by a story that I was told on several occasions. The tale in question concerned a woman who, in the aftermath of the Great Nordic Wars in the beginning of the 18th century, managed to deter the entire Russian navy.

The story went something like this: a woman was out fishing when she caught sight of Russian galleys approaching on the horizon. She feared that the force was heading for Sandön on a punitive expedition. Sweden was at the time under pressure to make concessions in the peace negotiations with Russia. In order to compel Sweden to accept its demands, Russia had dispatched an expeditionary force to burn down the settlements dotting the Stockholm archipelago. Sandön, located precariously at the very edge of the archipelago was vulnerable to an attack. The clever woman who recognized the looming threat hurried back to Sandön. There she instructed all the islanders to raise as many flags they could muster. Mistaking the fluttering flags from a distance for the Swedish Royal Navy, the Russians changed route and Sandön was spared from being sacked and burned.

That was the story – a wandering local consolation story of ingenuity and bravery in exchange for the loss of an empire. For some reason, the story has stayed with me: the plucky woman who could out-row the Russian galleys, the shrewd bluff and the evasion of an impending doom. Perhaps a part of why it lingered was that when I asked if it was true, people hesitated and changed the subject. (It is not true, at least not for Sandön. As it turns out several islands have a similar story. It is possible that something along these lines took place on Möja, another island.)

However, the dubious authenticity of the story did not seem to change the astute behavior of its main character. She remains cunning and decisive to this day. This is the strength of stories. They do not need to be factually true in order to convey a conditional truth that still can be instructive and evocative. This was hardly something I seriously considered as a nine-year-old boy – the summers in Sweden are much too short for that. However, it was something that I was reminded of when I started to think more systematically about deterrence and game theory, which (as it happens) is the subject of this thesis.

As an operations analyst working for the Swedish navy and the EU operation Atalanta during the end of 2000's and early 2010's, I tried to appraise to what extent the Somali piracy was deterred by the presence of the naval task forces. The more I tried to answer the question, the further I seemed to drift away

from a reasonable answer. It was not clear which data could be used in order to measure the level of deterrence. More importantly, what was a suitable metric for deterrence in this context? Deterrence is a central concept in military strategy, but it is mostly used to make sense of defense strategies on a national level, often associated with the nuclear strategies of the Cold War. In a military operation, the concept still exists, but is rarely operationalized. Game theory has the notable advantage of offering such a metric. It helps to cut through the noise and hone in on what really is the central parts of deterrence. It is therefore a go-to method for many operational analysts and perhaps even more so for economists to make sense of complex behavior. However, when one starts to think about how much sense game theory actually makes, this in turn poses new questions in terms of justifiable assumptions, human behavior, rationality, meaning and interpretation. This thesis is about these two interdependent subjects, deterrence and the scientific appraisal of its value when studied through the lens of game theory.

Stockholm, December 2021

Acknowledgments

First and foremost, I would like to thank my main supervisor, Professor Till Grüne-Yanoff, Division of Philosophy, The Royal Institute of Technology (KTH). Professor Till Grüne-Yanoff's insights into economics and decision theory as well as his general advice in the art of crafting scholarly writing has been invaluable to me. I am indebted to his time and intelligence.

Furthermore, I would like to extend my gratitude to Jan Ångström, my supervisor and Professor at the Swedish Defence University (FHS), for funding this doctoral project. Jan has been crucial in making my research relevant from a War Studies perspective.

I would also like to thank the members of the faculty at the Division of Philosophy at KTH for the general development of the articles and introduction of this thesis. In particular, Sven Ove Hansson, Tor Sandqvist and John Cantwell's comments and suggestions have been important for the evolution of this thesis. In addition, I would also like to thank Jesper Jerkert, Anna Wedin and Maria Nordström. In particular, I would like to thank Robert Frisk, fellow PhD student, who like me has divided his time between KTH and FHS and provided invaluable support.

As the faculty at KTH provided an atmosphere conducive to intellectual growth, so has the faculty at FHS. In particular, Jerker Widén, Patrik Hulterstöm, and Linda Johansson have been instrumental in the development of the thesis. In addition to these prominent individuals I would like to thank Anders Enström, former Head of the Naval Operations section at FHS and my other colleagues who have helped me hold the course throughout the work, provided technical advice as well as moral support.

It is at times challenging to write a PhD in a philosophy department, all the while trying to remain relevant to a subject such as war studies. I am therefore grateful for how the two faculties at KTH and FHS not only accepted but also supported my work.

I am thankful to Erik Mohlin, Lund University, who has been a faithful sparring partner on the scopes and limits of game theoretic analysis, but above all a good friend with learned advice and kind encouragement. On the same note, I would like to thank Johan Egerkrans, lifelong friend, author and illustrator who put pen to paper and made my whimsical idea for a cover into something better than I ever could have envisioned.

I would also like to thank Peter Wikman at the Swedish Defense Research Agency (FOI), who provided valuable advice and insight about prospect theory.

The articles presented in this thesis would not be what they are without the comments and suggestions provided by a number of referees, whose names I do

not know. This is as it should be, but if they ever get around to reading these lines, I hope that they too feel included in my appreciation.

The Swedish Royal Navy, in particular Jan Thörnqvist, Anders Larsson and Mathias Jansson were gracious in enabling the research presented in article IV as were the officers, sailors and soldiers serving on board HSwMS Carlskrona (ME02 and ME03).

Finally, I would like to thank my wife Antje and my daughters Elsa and Siri. It is a privilege to be surrounded by such intelligence, grace and curiosity.

While the author is indebted to many, all mistakes and shortcomings rest with him alone.

Introduction

1 A Game of Deterrence

> *While he stood there thinking, Achilles, peer of Ares, approached, the plumes of his helmet nodding, brandishing the mighty spear of Pelian ash in his right hand, high above his shoulder, his bronze armour blazing like fire or the rising sun. Now Hector was gripped by fear and, trembling at the sight of him, afraid to stand his ground by the gate, set off running.*
>
> *The Iliad, Book XXII – Achilles chases Hector round the wall.*

Most people have probably thought about deterrence in one form of the other: from musings about keeping intruders out with a burglar-alarm, or threatening a disobedient child with some sort of drastic consequence in order to alter its behavior, to heading off another car at a crossing – will you or it swerve? If we look beyond trivialities of everyday life, we find deterrence all around us: the wasp-mimicking markings on the hover-fly to ward off predators, businesses cutting prices to dissuade competitors from entering a market, and the positioning of nuclear ballistic missile submarines guaranteeing a second strike against a would-be aggressor. The mighty Hector loses his usual brave demeanor and bolts upon facing the greatest of heroes, Achilles. At heart, deterrence is a behavioral strategy for survival by altering the behavior of an opponent by threat. It is the will to avoid violent confrontation by threatening with violence – be it based on a true consequence or a mere bluff.

As a national defense strategy or a military stratagem, deterrence quickly becomes complex. Everything from planning and procuring military capabilities to positioning of military units and executing maneuvers become intricate parts of the phenomenon of deterrence. Add to that the deliberations, incentives, opportunities and decisions involved, and deterrence becomes nothing short of a multifaceted amalgam of human behavior and technology. Game theory has the ability to hone in on the basic strategic dynamic between actors, their strategies and possible outcomes. It recognizes deterrence on its most basic level, where the

actions of one actor are dependent upon the strategy chosen by the other, and vice versa. It is not strange that a method that cuts out everything that obfuscates the dynamics of deterrence becomes popular – for scientists and policy-makers alike. Given its popularity, the focus on game theoretic modelling has meant that the model is put front and center in the scientific discussions about deterrence. Theorizing about deterrence is often synonymous with theorizing about game theoretic models of it.

In the wake of the Russian annexation of Crimea, the Georgian War, heightened tensions between the US and China, as well as the risk of WMD proliferation, many countries are revising deterrence strategies originally devised during the Cold War. Amongst defence organizations such as NATO, there is a realization that whereas some things undoubtedly have changed since then, other aspects might still be comparable to the Cold War dynamics (see for instance Krepinevich 2019). So how should deterrence in the 21st century be approached? Are the old models still relevant? While the world may be adjusting to a new security political environment, the deterrence researchers have not been idle. New models and insights regarding human behaviour have changed some of the conclusions suggested by the Cold War modelling. But how can these new ideas be applied? Should they be applied?

While both the world and the research community have moved on since the time of the Cold War, some issues remain the same: the model's ability to represent, its relationship to theory and the evidence for a particular model are reoccurring topics. The main contribution of this thesis is, beside the research articles themselves, to analyse how we can approach these topics by considering them from a philosophical perspective. The question asked is: what is the game theoretic model's contribution to the study of deterrence? By answering that question, the game theoretic models used in deterrence research can be introduced to the reader and the philosophical issues associated with game theoretic modelling can be presented. The thesis' overarching aim is to investigate the dynamics of deterrence when studied with the aid of models, the model's relationship to theory and to evidence.

The research on deterrence is frequently model-driven; however, it is a scientific choice to use models – a choice not everyone has made. There is a rich research field concerned with deterrence that is not focused on modelling. Several scholars use a more informal approach: Herman Kahn's tome *Thermonuclear War* (1960), Alexander George and Richard Smoke's *Deterrence in American Foreign Policy* (1974), Glenn Snyder's *Deterrence and Defense* (1961), and, of course, Bernhard Brodie's *Strategy in the Missile Age* (1959), are all well-known accounts of deterrence theorizing. Later additions such as Lawrence Freedman's *Deterrence* (2004) and Patrick Morgan's *Deterrence Now* (2003) are also informal, qualitative

well-recognized discussions on deterrence. While not concerned with deterrence modelling *per se* most of these contributions refer to the model-driven literature and frequently borrow concepts and distinctions from it. Just as this tradition draws inspiration from the model centered literature, the reverse is also true. The model-driven approach often borrows concepts or latches on to queries from the qualitative research tradition. Thomas Schelling's famous *The Strategy of Conflict* (1960) which models various deterrence related situations is in many ways a response to issues discussed at the time of his writing. So, while modelling of deterrence in many ways can be considered a core of the research on deterrence, other traditions do play a role and the interaction between the different traditions has clearly been fruitful. This is especially clear with the quantitative research branch concerned with deterrence. Paul Huth's *Extended Deterrence and the Prevention of War* (1988) relies on both the formal and the qualitative distinctions in order to develop a statistical analysis of the phenomenon. While this thesis traces the model focused tradition, a separation of the different traditions from each other is not possible since it would miss the interactive nature of deterrence research between the different traditions.

The thesis is focused on the model-driven part of deterrence research. It unfolds in the following manner: the thesis introduction is divided into five sections. First the game theoretic model's representational qualities are discussed from a philosophical perspective. Central concepts and distinctions are considered and framed in the context of the models used to study deterrence. These insights are then carried over to the next section where the theories most prevalent in deterrence are presented. These theories' relationship to the models are mapped and discussed from a philosophical perspective. Next, issues concerning statistical and experimental evidence follows, before some general conclusions are presented. In the last section the three aspects of model, theory and evidence are discussed in order to take stock of the scientific status of deterrence research. Four longer abstracts of the research articles conclude the introductory part. Then the four research articles are presented in full.

Military power and armed conflict are challenging to make sense of: humans, metal and technology come together; sometimes in quick engagements, sometimes over long-time spans. Systematic research is a central process in order to distinguish patterns, separate causes from effects and theorize about the mechanisms involved. Game theoretic analysis of deterrence is a good example of how this can be done – and much has been accomplished. Knowing the scientific techniques is a first step towards an understanding of war and conflict. However, there is an inherit danger if the methods and concepts used in the sciences are taken for granted and their application goes without questioning. The economic and social values associated with national defense and military operations are

immense, where the former may define large part of society. False assumptions, misunderstood or misapplied concepts, will at best only carry enormous economic costs when applied, at worst they may rescind a nation altogether. It is therefore imperative to get the analysis of deterrence right. This starts with the assumptions and concepts that define it. The thesis illuminates some of the foundational issues regarding deterrence research, while also contributing to the subject matter itself.

2 Reality and Representation

While deterrence itself has received little attention so far, modelling of economic behavior, which essentially is what deterrence modelling is concerned with, has been treated quite extensively in the philosophy of science. So, what are the basic functions of the model? What type of knowledge do we attain from studying a phenomenon through a scientific model? On the most basic level the idea of the model is to identify the relevant structures of the phenomenon one wants to study. Philosophers of science tend to view the model-world relationship as a type of analogy (Hesse 1966; Giere 1999; Sugden 2000). An analogical relationship in this context means that the model (M) maps certain aspects of the target (T) thereby representing certain aspects of the target; studying the model thus aims at supporting the study of the target. Central is therefore to understand the M–T relationship, the analogy. Mary Hesse offers one of the more nuanced accounts which treats the analogy as consisting of two parts. The first part is the relation between the various aspects that exists *within* the object of reference (the model) and the various aspects in the object under investigation (the target). The second part of the analogy is the correspondence *between* the aspects in the model and the aspects in the target (Hesse 1966, 59).

The models in deterrence research are game theoretical models. They are models that try to depict the behaviour of players. It is generally assumed that the players try to maximize what they possibly can gain (or minimize what they can lose). One of the best-known deterrence models, the game of Chicken, is a case in point and illustrative of the analogical relationship between model and reality (the game Chicken figures in articles I–III. For a thorough presentation see article I). Hesse's division of the analogy can be applied to better understand what the model does and the relationship between the model and the target. The relationship *within* the model, Chicken, depicts two players who each have (the same) two strategies. Both players have a strategy to Cooperate and a strategy to Defect. The combination of the players' strategies results in four possible outcomes, represented by a 2×2 matrix: the *Status Quo* outcome, when both

players play *cooperate*, *Conflict*, when both players play *defect* and two *Advantage/Disadvantage* outcomes where the player who plays *defect* get the Advantage and the player who plays cooperate gets a Disadvantage. Each player has an order of preferences over the outcomes where they prefer to get an advantage over the other player, which they prefer to the Status Quo, which in turn is preferred to getting a disadvantage, which still is preferred over conflict (i.e., Conflict being the worst outcome for either player). This is the relationship within the model.

The relationship *within* the target, for example particular episodes of the Cold War, consists of two main actors the East and the West. Both of the two actors tried to attain an advantage over the other without being taken advantage of, or bringing about a full nuclear exchange. Broadly speaking, this is the relationship within model and target, respectively.

The analogical relationship *between* model and target is mapped out by exploring the positive qualities, the similarities, and the negative qualities, the dissimilarities. One similarity is that the number of players in the game corresponds to the number of the main stake holders during the Cold War. Further similarities concern the potential strategies for the players which reflect the two major choices, and the outcomes which mirror what possibly could happen in such a scenario. The dissimilarities, the negative parts of the analogy, are of course many more: the details of the capabilities, the positioning of military assets, the posturing of units as well as aspects such as the actual deliberations of decision makers, coordination between allies, etc.

Hesse's analysis enables a delineation between the internal relationships within the model, the internal relationships of the target and the relationship between the model and the target. An important aspect of the analogical relationship is the so-called "neutral part" of the analogy, where it is unknown whether model and target are similar or dissimilar. In Chicken a neutral part of the analogy is what happens if both players defect, i.e., that they play defect in the model, and the rivaling sides of the Cold War are readying their missile silos in preparation for a full nuclear assault. What would this mean? Is the model's prediction of nuclear war correct? Perhaps in such a contingency, both parties would be deterred? It is this neutral part of the analogy that opens up for questions and hypothetical reasoning. The analogical structure between model and target enables prediction, formulation of hypotheses and the prompting of questions. This aspect is particularly noticeable in the quantitative research on deterrence, where the research design often centers on what a given model predicts, and then compare the results from the statistical analysis with the model's prediction (see, for instance, Huth and Russett 1984; Quackenbush 2010).

2.1 Similarity

A recurring question concerns what similarities are needed in the representation in order for the model to provide a useful representation of the reality it aims to depict. Some models have an isomorphic relationship to its target, i.e., the correspondence between model and target is in a one-to-one relationship, like the blue print of a diesel combustion engine and actual diesel engine, where each part of the blueprint has a corresponding part in the actual engine. The game theoretic model is different since it is a representation where only some aspects carry some type of similarity; the model idealizes the representation of the intended target (Morgan and Knuutila 2012). The question will therefore ultimately turn on what aspects should go into the model in order for it to represent.

Scientists typically model a phenomenon by considering a number of aspects, actively choosing among them in order to construct the desired analogy. The aspects that are chosen are intended to be of relevance and similar to its target (Giere 2004; Mäki 2009). One can imagine how Chicken could misrepresent if we added a player or changed the preferences order. The critique of Chicken typically focuses on the lack of similarity to the Cold War, such as the structure of the game and the preference orders of the players (Powell 1990; Zagare and Kilgour 2000; Zagare 2018). If the model is detached from its target, then the analogy falters and inferences from the model will not have sufficient bearing on the target. Hence, according to some critics, the right kind of similarity is missing to make Chicken an acceptable representation of the Cold War.

When the similarity between Chicken and the realities of the Cold War has been claimed as inadequate, defenders of the game have maintained that the similarity is sufficient for a general type of representation. Further, it has been argued, more nuanced accounts that try to map the target of the Cold War more closely, needlessly complicate the dynamics, thereby obscuring the most central aspects (Walt 1999; Lawson 2013). Philosophically, one can take this one step further and ask to what extent focus on similarity is required at all. In an often-cited article Robert Sugden maintains that the importance of a model is not whether it represents its target *per se*, but rather whether it conveys a convincing image of a situation, what Sugden calls a "credible world" (Sugden 2000). To Sugden, the strength of the model centers around what it can tell us about a phenomenon; the similarity to the actual target is of subordinate value. Applied to Chicken we can think of it as a generic game, which may not actually have taken place, but represents a convincing picture of what the essence of mutual deterrence is. From this general picture, we can infer various conclusions. This view is also more in line with how early deterrence scholars like Schelling and Kahn seem to relate to a game like Chicken, as an inspiration and point of departure,

rather than as a fixed object dictating the strategic dynamics of the Cold War (Schelling 1960; Kahn 1960).

2.2 Purpose

Given the partly differing ideals of focusing either on the credibility of a model or aspects that are similar between model and target, it is clear that the intent of the scientist doing the modelling matters. Giere therefore suggests that the *M–T* relationship is replaced with "*S uses M to represent T for purposes P*"; where *S* = the scientist (or group of scientists). The difference being that now the scientist(s) and the purpose of creating the model are made explicit (Giere 2004, 743). For instance, a model that is made to explore credible threats is likely to focus on other aspects than a model that aims to depict deterrence by risk-taking. Hence, there is a non-trivial element of any model where the modeler's interest is what dictates what goes into the model and what is left out of it. On one level, this may seem obvious; however as pointed out by Clarke and Primo (2007), it is a recurring phenomenon within the social sciences that models are put to use and evaluated in different domains than they originally were intended for. The practice of using models in types of studies they were not intended for is in perfectly good order, however any critique of a model needs to take the original purpose of the model into consideration, if it does not it, the critique is out of place (Clarke and Primo 2007).

The game theoretic model is a representational model that idealizes certain aspects of its intended target. Issues such as similarity and credibility can be important to consider when appraising a model's representational qualities. However, such issues must always take the purpose of a given model into consideration.

The article in this thesis that most explicitly refers to the philosophical aspects of what the model's role is in a given context is article III. As mentioned, a reoccurring discussion in deterrence research is which model best represents the phenomenon of deterrence. Article III aims to settle this question by proposing a method for comparing different game theoretic models with each other. It considers three aspects for comparison: (i) how a model relates to its *target*, (ii) the *criteria*, i.e., how the model is constructed, and (iii) what the *purpose* of the model is. With these distinctions, it can be shown that the three models that tend to come up in the discussion about deterrence representation (Chicken, Entry Deterrence and Perfect Mutual Deterrence) relate to (partly) different targets, rely on different criteria and are designed with different purposes. This underscores how dependent a model is on several factors, which, if neglected, can lead to misunderstandings and confusion. The analogical relationship is fragile. There are only certain aspects that have bearing, and it is only such aspects that should

be represented in the model. If we misinterpret or misunderstand the model, then it is highly likely that we draw the wrong conclusion about the target – i.e. about reality.

2.3 Model Deconstruction

Before turning to what issues such as similarity, credibility and purpose implies for a model's relation to theory, we need to know what type of idealization a game theoretic model is concerned with. Grüne-Yanoff and Schweinzer (2008) suggest that the game theoretic model can be viewed as consisting of three parts or layers: *structure*, *narrative* and *theory proper*. The structure is the matrix, the strategic form game or the extensive form game. These forms, the structures, are what we often first think of when we think of game theory. The structure displays the players, their potential strategies and possible outcomes. The narrative is the story that accompanies the structure. In a game such as "Battle of the sexes", it is the story about a man and a woman who try to coordinate whether to meet at the opera or at a boxing match; in the game "Stag hunt", it is the story about hunters trying to cooperate. The narrative provides the structure with a meaning. Theory proper consists of the solution concepts that direct how we understand the consequences of playing the game and the existence proofs. It is theory proper that stipulates the Nash equilibrium with pure and mixed strategies (Grüne-Yanoff and Schweinzer 2008).

Connecting this division to the analogy, the structure and narrative of the model are complementary accounts of the target: they depict the target in different ways. The theory proper, on the other hand, is not necessarily part of this depiction; its function is to stipulate how we can conclude what is going to happen and infer this to the target. It is the theory proper that forms the rules of how we are to relate to the neutral part of the analogy.[1] In Chicken, theory proper dictates that there are only two outcomes that are possible with pure strategies; one defects and the other cooperates. It may be that this will turn out to be the wrong prediction for a given situation, but with the theory proper's stipulations the defect/cooperate outcomes are the only two theoretically possible. The division into structure, narrative and theory proper is instructive as it reveals the vital parts of what a game theoretic model is doing: it shows the dynamics of

[1] Structure, narrative and theory proper are used in article II to compare Russell's (1959) original account of Chicken with the game theoretic version of Chicken. Russell's version and the game theoretic version have identical structure and narrative, but consequence of playing the game differs between the two accounts. Russell's conclusion is that such a game is dangerous to play since mistakes and misunderstandings are likely to put both players on a collision course. The game theoretic version, with its theory proper, is that when analyzed with pure strategies one player always swerves, with mixed strategies collision is a possibility.

a social interaction with an accompanying story and defines a concept of how we are to draw our conclusions.

Theory proper is what sets game theory apart from other possible methods of relating to social interaction. What we may learn from a game is dictated by the theory proper. This means that when deterrence interaction is analyzed with game theory, it is the theory proper that defines deterrence behavior. This also means that the theory proper of game theory becomes a part of the theory proper of deterrence theory. As discussed in article II, this can be considered a constraint when analyzing deterrence. Schelling therefore departs from the theory proper, but retains the structure, Russell never applies it. The issue of how much a deterrence theory should adhere to game theoretic analysis is a recurring topic in deterrence research (sees particularly Field 2014, but also articles I and II).

If we make a theory of deterrence dependent on game theoretic reasoning, this clearly enables sophisticated analysis, but it might also constrain a theory to the confines of game theory. Likewise, ignoring or downplaying game theoretic analysis could also seriously limit an attempt to create a theory. In game theory, we attain a type of extreme analysis where the behavior of players optimizing their decisions is analyzed. Ignoring such contingencies to only analyze actual behavior or recounting historic events would miss out on what could happen if players optimized. Article I analyzes what happens to a deterrence theory if we exchange the standard game theoretic representation with an alternative, a prospect theoretical function (see also Carlson and Dacey 2006; Butler 2007).

What value we place on the models that figure in deterrence research in relation to a deterrence theory is therefore not altogether straightforward; similarities between model and target, the intended purpose as well as assumptions such as rationality will always play a role in modelling a complex behavior such as deterrence. The status of the model matters. If we believe its representational qualities are representative of deterrence dynamics, the model is an essential part in a theory of deterrence. As a consequence of focusing on a model for a theory, validity of the model's predictions needs to be appraised. However, if we use and relate to the model as a generator of hypothesis or as a heuristic for discussing deterrence, then its value for theory will be less crucial. In such cases, validation is not necessarily needed. If the model is used as a type of inspiration for a general theoretical discussion, model validation can be dispensed with altogether.

3 Theories of Deterrence

The previous section made the case that the contributions of a given model varies and depends on intent as much as on the content. This will in turn have implications for the formulation of a theory. What does a game theoretic model imply for a theory of deterrence? In the literature on deterrence, references are regularly made to a deterrence theory, but it is often left unspecified what a theory of deterrence actually is or consists of. This is so because one can identify several theories or versions of theories of deterrence. Hence, in order to appraise the model–theory relation we need to discuss what this relationship consist of and how it has developed.

3.1 Origins

Historically, theorizing about deterrence can be traced back to the Vegetius dictum *"si vis pace para bellum"* – if you want peace prepare for war – but as the Homer quotation in the first section indicates it had been recognized earlier by the Greeks.[2] Given that deterrence also is found in nature, one can conclude that it is a very basic survival strategy and theorizing about it has been an ongoing human endeavor. An issue with deterrence has always been that a defender needs to make its threat believable for a potential aggressor in order for the latter to refrain from attacking. Just as this is true for the hoover-flies, Romans and Greeks, it still holds true today with regard to conventional deterrence. With the development of nuclear weapons such issues became supercharged: now both actors need to show resolve about attacking if attacked.

This, in and of itself, was not new. Being both challenger and defender at the same time has been possible also in conventional situations, like the circumstances leading up to World War I. However, the destructive power of a massive nuclear engagement is something different (Wirtz 2018). War has always meant destruction, but with nuclear weapons, the destruction went from limited to a specific territory, to challenging the existence of human civilization. Eisenhower's New Look doctrine clearly stipulated US reliance on its nuclear capability and the USSR strategy was in many aspects similar. Since a Ragnarok-scenario by most is viewed as the least preferred outcome, the issue becomes how one can make a threat that also would eradicate oneself credible? With such a strong doctrine, other issues arose too: is it credible to execute a nuclear war because a coup d'état in a faraway land was sponsored by one's enemy? The problem of

[2] The full quotation from Vegetius reads: "He, therefore, who desires peace, should prepare for war. He who aspires to victory, should spare no pains to form his soldiers. And he who hopes for success, should fight on principle, not chance. No one dares to offend or insult a power of known superiority in action." *De Re Militari*, Book III (Vegetius 2012).

credibly threatening nuclear war in every contingency gradually became apparent with the US doctrine during the late 1950's. A central question was: whether such a relationship is reciprocal and if so, is it stable? Was the world constantly on the brink of exterminating itself? Several thinkers both in- and outside the military high commands definitely thought so (George and Smoke 1974, 21; Kaplan 1983).

To help settle such questions, deterrence strategists turned to game theory and from its inception, deterrence has been closely linked to game theory. Roger Myerson (1991) suggests that the close relationship between game theory and deterrence is due to the fact that many scholars were involved both with the Manhattan project and with the development of the US nuclear capability. Pillaring over all of the many men and women who developed the weapon itself and the policies that came with it is of course John von Neumann, founder of modern game theory (Kaplan 1983; Field 2014; Zagare 2018). Game theory and deterrence became interlinked.[3] Even critics like Bertrand Russell used games in their attempts to lay bare the dangers of nuclear competition (Russell 1959).

3.2 Classic Deterrence Theory

Zagare, Kilgour and Quackenbush suggest a thorough account of distinctions between different deterrence theories, focusing on their relationship to modelling (Zagare and Kilgour 2000; Zagare and Quckenbush 2016). The trio identifies three distinct theories that they associate with three different types of underlying assumptions. These three schools of thought can be traced back to the analysis of the strategic dynamics of the Cold War. The first (i) account is what they call the *structuralist account* of deterrence. This account's attention is placed on the structures of the relations between nation states and the costs associated with a full nuclear exchange between the US and the USSR. While the focus of the theory is the Cold War's power structure, it also theorizes about other power relations prior to the development of nuclear arms.

The structuralist view is closely associated with another theoretical paradigm known as *balance of power*. The balance of power theory alludes to the idea that if two actors have an equal level of capabilities, war will not occur – the power balance makes conflict unattractive to both sides, i.e., they deter one another. Some scholars, like Intriligator and Brito, also incorporate the notion that the level of capability must be relatively high for deterrence to occur (Intriligator and Brito 1984). According to this slightly more nuanced account; deterrence fails if there is an asymmetry in power or if the capability of both actors is too low.

Zagare, Kilgour and Quackenbush point out that structuralist deterrence

[3] Two Nobel laureates have received their accolades in part for their research on game theory and deterrence: Reinhard Selten in 1994 and Thomas Schelling in 2005.

theory is not *per se* connected to game theoretical modelling and while it may allude to actors as rational, it is often a different type of rationality than the standard assumption in economics and game theory.[4] When scholars rely on the standard definition of rationality it is the second (ii) account of deterrence theory, the so-called *decision-theoretic* account. This branch is related to the structuralist deterrence theory but stresses game theoretic modelling. It adheres to the same basic idea of balance of power but with a game theoretic underpinning. It is in this context we find early deterrence models such as Chicken. The structural deterrence theory and the decision–theoretic deterrence theory are loosely roped together as the classic deterrence theory. In this tradition we find Schelling, Kahn, Jervis and most of the other deterrence scholars from the 1960's and onwards.

Classic deterrence theory is largely dominated by the discussion about two deterrence models, Chicken and the somewhat later Entry Deterrence (Zagare and Kilgour 2000; Quackenbush 2011; Sörenson 2019). Schelling (1960), Kahn (1960), Jervis (1973) and Snyder (1976) all discuss and relate to Chicken. However, the model is not a deduction from a greater theoretical construction; rather it is used as inspiration or departure for a theoretical discussion. Kahn considers it as carrying a dangerous reminder of what could ensue (Kahn 1960, 291). Jervis uses it to discuss how nation states tend not to conform to the logic of this behavior (Jervis 1973). Similarly, Snyder discusses it before developing other psychological insights about deterrence (Snyder 1973). The models seem to have the function of informing theory, rather than defining it. The models and reasoning about model-assumptions are used in order to explore the potential for a theory and what would make such a theory. A good example is Schelling's idea of a "trip-wire". Schelling points out that the main task for the US marines stationed in Berlin during the Cold War was not as a fighting unit, but as a strategic hindrance, a trip-wire, for an advance from USSR into West Berlin. If the Soviet Union had made a push the marines would have fought and died, thereby committing US forces to join its Western European allies on the battlefield (Schelling 1966, 105). To these early deterrence scholars, the question whether Chicken or Entry deterrence were similar to the Cold War seems not to be the main issue. Rather the games are used to imply a type of dynamics that was important to take note of. To paraphrase Sugden, the games seem to function as "credible worlds of deterrence". To the extent that one can discern a distinct view amongst the

[4] Also here Zagare and Kilgour assist us with a distinction, namely what they term as procedural rationality and instrumental rationality. The former is identified with agents that "begin with establishing an accurate understanding of their environment and by eliminating all conceptual impediments to sound choice" the actors also factor in their own biases and other shortcomings before making their decision on a strategy (Zagare and Kilgour 2000, 39). Instrumental rationality is the standard rationality definition from Luce and Raiffa (Luce and Raiffa 1957, 50) and discussed in the previous section.

deterrence theorists of this time, they seem to rely on the models as mediators of theory. The models inform, inspire and explain what a deterrence theory could potentially become. However, also the game theoretic underpinnings are at times disregarded in order to follow a particular line of thought. Article II discusses Schelling's departure from the theory proper of game theory in order to pursue a different type of logic he identified as important. What is present is the strong discussion of the narrative. Hence, the models are important, but as points of departure that, together with other aspects, should be regarded as foundational for a theory. Consequently, these first deterrence scholars seem not to view the models, Chicken or Entry Deterrence, as synonymous with theory, nor do they seem to see the model as a necessary a part of the theory, but as a mediator of theoretical insights.

3.3 Developments

The potential fragility of the Cold War changed with the development of second-strike capabilities. With nuclear ballistic missiles submarines hidden under the polar ice cap, it became impossible to strike the opponent without these vessels being able to deliver a counter strike. Now an aggressor could no longer preempt an attack by striking first, since it would still be exposed to a retaliatory volley from the opponent. Running parallel to the development of the second-strike capabilities was the realization that the nuclear threat was difficult to make credible when a threat was on the lower end of the conflict spectrum. The Eisenhower era New Look strategy was replaced by the Kennedy administration's strategy of flexible response. Now conventional threats were to be matched and trumped by one's own conventional forces. In this way other types of contingencies could be handled without having to escalate to the nuclear level (see Schelling 1960 about surprise attack; George and Smoke 1974, Keeny and Panofsky 1981, Powell 1990, regarding second strike strategies and flexible response). Besides necessitating massive increases in defense budgets across the board, these two developments also prompted new research questions, primarily, whether this new development also meant a more stable relationship between the East and the West.

An issue for the deterrence theorists analyzing the Cold War was that if the new relationship in fact was relatively stable, then the model Chicken misrepresents. In Chicken, the Nash equilibrium states that either of the two players gets an advantage. This is not how the Cold War seemed to have played out. To the extent an actor "defected", it was a rare occurrence. At minimum, the game is inadequate. Powell (1990) is clear: with the devolvement of second-strike capabilities, Chicken does not capture the strategic dynamics since a defection by one player could be countered with a defection by its opponent (Powell 1990,

35). Instead, Powell remodels the intuition about threats that leave something to chance. A credible threat, in Powell's analysis, is a threat that escalates the probability of conflict to such an extent that an opponent is deterred from taking the next step in the escalatory ladder since this step will push both players beyond the brink into a nuclear exchange (Powell 1990).

Also, Zagare and Kilgour formulate their critique to the early theorizing by departing from Chicken (Zagare and Kilgour 1993; 2000). It is their own analysis of deterrence that is the third (iii) account of deterrence theory: the so-called Perfect Deterrence. Unlike classic deterrence theory, Perfect Deterrence theory departs from a basic game-form where deterrence is achieved when the players not only have sufficient capability, but also have credible threats. Deterrence is thereby attained when the players are credible. From this basic game-form they model asymmetric, mutual and extended deterrence (Zagare and Kilgour 2000). In Perfect Deterrence the models are the foundation for the theory and a problem with the model would directly affect their theory. When the models become this central it also becomes important to understand if they represent, not just a credible possible world, but a real existing world.

The differing ideals guiding the modeling practices in the later development take a different tack when it comes to the actual modelling as well as the models' relationship to theory. Concerning the modeling practices, Zagare, Kilgour and Powell represent a tradition where they not only suggest games that in some form capture deterrence interaction, they also formulate a theory to go with it. In doing so they depart from the narrative driven part of the analogy: hooligans and competing firms are replaced with the historical records to exemplify their respective dynamics. The rudimentary constructions of the original games are replaced with more refined games. A more nuanced similarity between target and model is clearly a guiding aspect. This emphasis on model-theory coherence is not confined to deterrence research. A similar approach can be found in Woosang and Buena de Mosquita (1995) as well as in Slancthev's (2012) writings on conflict and coercion; here, too, model and theory are closely intertwined.

3.4 The Theory–Model Relationship

How should we understand the various deterrence theories reliance on modelling? Scientific theory is often divided into three parts; the syntax, the semantics and the practices. If emphasis is placed on the abstract grammatical structure of theory, the syntax, then theory is a set of sentences. These propositions' reducibility into axioms is therefore central. The syntactic view understands the model as a construction based on sentences that are deduced from the theory (da Costa and French 1990). Game theory can be reduced and understood in this syntactic

way. The decision theoretic axioms provide the foundation from which the von Neumann–Morgenstern representation theorem function can be deduced.[5] By this account a given model is of subordinate value since the theory proper defines the axioms and representation theorem from which any model is an extension. However, it is less clear how such account of theory can be related to a deterrence theory.

As we have seen, the use of models is central for deterrence research, but axiomatization is not something deterrence scholars are concerned with. If anything, deterrence theorists tend to relate to theory and model in different ways. Sometimes the model relates the essential dynamics from which one can make predictions. This is how Schelling and Kahn seem to relate to game theoretic modelling. Weisberg (2013) discusses the model as independent from theory altogether and exemplifies with the Lotka-Volterra model of predator–prey interaction. The model is created independently of theories in biology. If models are viewed as such, they are in a sense substitute for an actual theory. This, on the other hand, seems not to be the case with deterrence. References are made to various empirical records of deterrence interaction; theorizing about deterrence, albeit without attempting axiomatization, is commonplace in the research literature (see for instance Morgan 2003). Models clearly matter in deterrence research, but so does theory. Consequently, the syntactic view does not seem to capture what is important to the deterrence scholars as axiomatization neither defines or directs the research.

If the syntactic view might be insufficient to describe how researchers tend to relate and regard their own research, one can instead focus on the semantics of a theory. In this case, the model becomes central. Such a view maintains that a scientific theory is the collection of the models within it. Therefore, coherence between model and theory is central (van Fraassen 1989). The model is thought to be closely related to the theory, i.e., "… the theory is entirely true if considered with respect to the model alone." (van Fraassen 1989, 218). In deterrence research, some models are based on theoretical insight about deterrence. Powell, Nalebuff, Zagare, Kilgour, Quackenbush, Langlois and Langlois all formulate their games with reference to earlier deterrence research. However, just as much as they relate to earlier theories, they use the models to explore a deterrence dynamic and then draw theoretical conclusions about these dynamics, thereby using the models to develop the theory. Particularly, Powell, Zagare and Kilgour seem to entertain an ideal akin to the semantic view, where there is a coherence between model and theory. Their developments are however motivated by deficiencies and problems with earlier deterrence accounts.

[5] Carnap, Reichenbach and Hempel are some of the more known proponents of this idea (Isaac 2005; Creath 1987).

A central problem in deterrence research is the issue of credibility. Both Powell's Nuclear Deterrence and Zagare and Kilgour's *Perfect Deterrence* are responses to how the problem of credibility can be solved. Deterrence is in essence a threat, but what is a threat? This question has occupied deterrence scholars for many decades. Schelling and Kahn proposed various ideas (see article II), such as threats as pre-commitment, where a player ties its hands to carrying out a particular threat (see also Slantchev 2012, who develops this idea). Alternatively, the idea of threats that leave something to chance is the issue of making the opponent unsure of the actions that the other player is willing to execute (Nalebuff 1986). Or irrational threats – the idea of behaving irrationally, thereby making an opponent afraid that the player is going to choose conflict by mistake (article II deals with this question extensively, but see particularly Schelling 1966; Field 2014, Zagare 2018).

Both the syntactic and the semantic views give descriptive accounts, but they are also normative accounts of what scientific research should strive for. However, they tend to overlook how science actually is conducted. An attempt that seeks to rectify this deficiency and one which takes a more holistic approach is the so-called pragmatic view. This school of thought views models, laws and axioms as important, but it also considers practices and problems within a given scientific field. What the scientists actually do and how they relate to their work is just as vital as the theories they are working with. If scientists tend to work, discuss and relate to their models more than to a theory, then this suggests that such practices should be given more attention (Suárez and Cartwright 2008). This attitude positions the pragmatic view as less normative driven than the semantic and syntactic views. If a problem like credibility has defined much of the deterrence research, then it has effectively shaped the direction of the research and explains why it is where it is today – and this fact needs to be accounted for when discussing the research field.

The pragmatic view is helpful in that it considers the scientific process. This enables examination of scientific developments that lie outside the scope of the syntactic view and the semantic view. It is a reasonable conclusion that Zagare, Kilgour and Powell's more model-centric versions of deterrence would not be where they are today had it not been for the early inadequate modelling of threats, which stimulated research that gave rise to more exact and stringent accounts. In this context, the pragmatic view offers more leeway than other perspectives in that it incorporates accounts that stretch beyond model-theory relation. Nancy Cartwright points out that a model that misrepresents can be just as important for the development of a research field in that it focuses the attention on the fact that something is wrong and needs rethinking (Cartwright 1983). The faulty model has no place in the syntactic or the semantic view. The normative aspect

of the pragmatic view is that when a practice or problem is deemed as central for a particular type of research, then this process or problem also ought to be given due attention. Aspects such as theory–model coherence are still relevant, but there are also other ideals to consider. Ideals are therefore in part relative to the research field they exist within and in part components of larger philosophical considerations.

While the distinction between the early deterrence theorists that viewed their models as mediators for theory and the later deterrence theorists that relate to their models as the core of theory, it is not clear if either of these perspectives is more reasonable than the other. If models are viewed as mediators, then the insights a model provides are optional: if a model fits a situation, it is of interest, if not, it can be dispensed with. However, if a model is thought to be the core of a theory its insights are not optional. Rather, verification whether a model actually represents the dynamics it aims to depict becomes crucial. Also, underlying assumptions become objects for examination, are they too strong? Too weak? Wrong? A foundational concept such as rationality is a case in point where deterrence scholars have turned to alternative assumptions that could improve deterrence representation. Such issues are less of a problem if one views the model as detached from the theory. As mentioned, Schelling disregarded the Nash equilibrium to pursue a different train of thought.

So far, focus has been on representation in the form of game models and their relationships to theory. However, both model and theory are subject to comparisons to the external world that they try to account for. Evidence for a model consists of information relating to the phenomenon it is aiming to depict or that a theory captures the basic causal mechanisms is vital, particularly if one views a model's similarity to a phenomenon as central.

4 Statistical Evidence

As mentioned in the introduction, the literature about deterrence does not only consist of discussions about deterrence modelling. A narrow but important strand in the literature is the quantitative approach of statistically studying deterrence. In economics, a common approach is to test a model once it is sufficiently developed. This is often done by means of statistical analysis, typically some type of regression, controlling for other relevant possible causes. In this manner, the external validity of the model can be appraised. Given the issue with the representational quality of some games, this could potentially settle the question which model best captures the strategic dynamic of a deterrence situation.

A statistical model differs from a game theoretic model. The game theoretic

model accounts for a type of theoretical dynamics. The statistical model has a different function. This model organizes data and information pertaining to a phenomenon that one is interested in. The organization and systemization are performed according to some type of sorting criteria. Patrick Suppes argues that statistical models are foundational for the empirical sciences since it is with this type of modelling that we can verify if our theoretical models represent the phenomenon in question (Suppes 1962, 252).[6] The question regarding which data that should be made a part of the statistical model and which data should be left out is a similar question to the what aspects that should be a part of a representational model. However, instead of causal aspects, issues pertaining to data selection are in the center of the deliberation. So, while the statistical model and the game theoretic model are different, the purpose of the study matters as much for both. The fitting of data and the selection criteria for the statistical model, taken together with the questions associated with the theoretical model, makes Ronald Giere (1999) conclude that science is largely a question about models. – There are models at the top, with the so-called models of theory and there are models at the bottom where we find models of data (Giere 1999, 42; see also Sörenson 2019 regarding examples in deterrence).

A central contribution to the deterrence literature, and of particular importance to the statistical research, is the discussion between Huth and Russett (1984) and Lebow and Stein (1990a) about conventional deterrence (see also Huth and Russett 1988; Lebow and Stein 1990b; as well as Lieberman 1994). Huth and Russett perform a statistical study of so-called *extended immediate deterrence*. *Immediate deterrence* is the deterrence behavior of state actors when a crisis already is underway. It is contrasted with *general deterrence* that concerns a state's overall deterrent relationship to potentially hostile states. *Extended deterrence* is when a state actor extends its threat to a third party to help deter a potential challenger. When only two actors are involved in a deterrent relationship, like the games previously discussed, it is referred to as direct deterrence (Morgan 1983). The discussion between Huth, Russett, Lebow and Stein is of interest for two reasons: it shows how difficult it is making a case selection of immediate deterrence (which still was viewed as easier than making a case selection for general deterrence) and it is a good example of how concepts and distinctions from various research traditions come together to solve challenging issues. Regarding the first issue, the case selection, the discussion treats how to separate false positives or type 1 errors, situations that seems to be cases of deterrence, but are not, from false negatives, type 2 errors – instances of deterrence that are

[6] Models like game theoretic models are sometimes referred to as models of theory since they essentially are accounts of causal relationships whereas statistical models are referred to as model of data.

rejected when they should have been included. The issue is coupled with the fact that it is difficult to know if an agent decides to forfeit a challenge because it is deterred or because it simply decides to do something else.[7] Further, Huth and Russett's original study departs from the assumptions deduced from a deterrence model. The model is not formally defined, but is a variation of Entry deterrence with variable payoffs depending on types of players (Huth and Rusett 1984, 500). Thus, statistical analysis ties what loosely could be called qualitative deterrence studies with formal model-driven approaches in order to create a viable research design.

Stephen Quackenbush (2010) moves the research forward when he studies direct general deterrence by departing from the Perfect Asymmetric Deterrence. Quackenbush uses the outcomes of the game as dependent variables. He then controls for balance of forces between the rivaling states, foreign policy, geographical proximity, democracy and peace years (Quackenbush 2010; 2011b). The study shows that the Perfect Asymmetric Deterrence game predicts the cases. This allows Quackenbush draw the conclusion that 1) the Perfect Asymmetric Deterrence game captures general deterrence between states and 2) credibility is central for explaining when deterrence succeeds.[8] Quackenbush then extends the analysis to incorporate extended general deterrence (Quackenbush 2011b).

Given that the Perfect Asymmetric Deterrence model seems to capture deterrence behavior – theoretically, as well as empirically – it is of interest to see if it can be applied in a similar but different domain. Most deterrence research is concerned with inter-state relations. However, deterrence is a defense technique that is not limited to the so-called strategic level. Military maneuvers on operational or tactical level can and do at times have the objective to deter an aggressor; either as a part of a larger engagement or as a solitaire maneuver. Article IV applies the Perfect Asymmetric Deterrence game to study if the naval intervention in Somali waters to deter the piracy was effective. The study uses a similar statistical model as the Quackenbush study and lets the outcomes be the appraised level of activity between pirates and naval units. It then controls for shipping's own level of security and the degree of turbulence in Somalia (Article IV, tables 2 and 3). The outcomes are then analyzed to appraise the equilibria suggested by the game model. The study shows that the naval operations were sufficiently credible to deter the Somali pirates in one area of operation (the Gulf of Aden), but lacked

[7] Morgan's original classification is not motivated by concerns of statistical case selection but as a classification aimed at better distinguish between different types of deterrence situations in general (Morgan 1983). However, the distinction is consistently used within the statistical research tradition as well as within the model-driven literature.

[8] Similar research has been done by Woosang and Buena de Mosquita who study crisis bargaining by employing a game model, coding the outcomes and then using multinomial regression analysis (Woosang and Bueno de Mosquita 1995).

credibility to deter the pirates from the other larger area of operation (the Somali Basin) (article IV, figure 4; tables 4 and 5). While the study is not designed to test the model, the Perfect Asymmetric Deterrence game provides a metric for classifying the possible outcomes from which the equilibria can be estimated, thereby enabling analysis of whether and to what extent the Somali pirates were deterred.

Suppes' contention that models of data are at the heart for empirical research as well as Giere's point that the scientific project mainly is a project concerned with modelling are certainly applicable to Huth and Russett's studies, Quackenbush's analyses and article IV. Crucial for all is the selection, organization and systematization of data: information regarding geography, levels of security and other types of data are discussed throughout. One can also note that several of the used variables in themselves are models of data. For instance, the geography variables that figure in the studies are themselves appraisals that are fitted into the models (see Quackenbush 2010; 2011b; article IV). The statistical models are then connected to the game theoretical models, models that in turn are built on assumptions regarding rationality. This is neither wrong nor strange, but it is a reminder that there is a fair number of assumptions and decisions regarding data that go into the research.

The many choices the researcher makes regarding data and type of statistical model is indicative of how purpose driven also the model of data is. Huth et al. defines and selects deterrence cases in order to analyze when the deterrence succeeds and when it fails. Similarly, Article IV appraises the level of deterrence in one specific case with the help of a game theoretic model and a statistical model. These two studies can be contrasted with Quackenbush's study that examines a game theoretic model's predictive power with the help of statistical analysis. The purpose of the Quackenbush study is different compared to the former two and as a consequence the statistical method is chosen to test the model (Quackenbush 2010). In the former cases the model is chosen to describe and analyze an empirical phenomenon. Hence, just as the choice of game model is informed by the study, so is the selection of data and the choice of statistical model. Thus, the purpose of study will play as much a role in the selection of the statistical model as it does with regards to the game theoretic model. What they are put in use for is what ultimately decides their potential contribution.

With more than fifty years of research leading up to the Perfect Asymmetric deterrence game, there is a certain degree of clarity that conventional deterrence can be represented with this model. Further, it is possible to use this game to make sense of deterrence on a lower than the strategic level of a conflict. The Somali piracy study in article IV suggests that also on a tactical level the model can be used to predict outcomes and to suggest equilibria. Hence, there are

strong arguments for believing that the Perfect Asymmetric deterrence model does capture a general deterrence dynamic that stretches beyond its originally intended area of use. While it may not be the model to end all other models, it is certainly a natural place to start for anyone who wants to understand asymmetric deterrence.

The issue with nuclear deterrence is thornier in terms of evidence than conventional deterrence. Quackenbush's study (2011b) does not discriminate between conventional and nuclear deterrence. However, since the conflict outcome has never occurred empirically in the case of nuclear deterrence the evidence is less compelling. This does not mean that nuclear deterrence cannot be studied, simply that the confidence must be viewed as lower.[9] This is arguably one reason for turning to game theory in the first place, since it offers a possibility of studying the strategic interaction theoretically in more detail and to reason about it in terms of counterfactual. The importance of analogical reasoning becomes quite apparent. There is, however, an alternative. If data on nation states' behavior are missing, turning to the original assumptions presents a substitute method.

5 Experimental Evidence

An alternative route to anchoring a model in reality is to turn to its theoretical underpinnings. An alleged shortcoming with game theory is the rationality assumption (see section 2). The assumption states that an agent will maximize its expected utility; however, it has been pointed out that people do not always behave in this manner, and when studied in experiments, results show that people tend to diverge from expected utility predictions (see, for instance, Camerer 2003 for an extensive treatment on the subject). If the idea is to realistically depict how actors behave, the rationality assumption is potentially too strong. As noted earlier, Russell's point with respect to Chicken is that it is a dangerous game to play precisely because people are not always rational. An assumption of rationality might therefore lead astray if it is not the realistic assumption to make.

If one changes the utility function to another type of representation, based on how people actually make choices, it potentially serves as a more realistic base for analyzing deterrence interaction. There are a number of alternatives to the standard von Neumann–Morgenstern function. From the point of view of modelling, the issues pertaining to experiments are essentially the same as with statistical models. The original model predicts a certain set of outcomes and the experiments show if and to what extent the data diverge from the model

[9] See for instance Narang (2012) who investigate different nuclear postures for effective deterrence and Rauchhuas (2009) who studies the nuclear peace hypothesis.

prediction.

The research program that perhaps has received the most attention is *prospect theory*. Daniel Kahneman and Amos Tversky, the researchers who proposed prospect theory, studied how people made monetary choices when faced with a binary option between a sure (or almost sure) gain or loss, and an option that involved a higher degree of uncertainty but also higher potential for a gain or loss. The way people made their choices suggested that when faced with a gain, people preferred the sure thing. However, when the choice concerned a loss, test-subjects were to a larger degree willing to take a risk (Kahneman and Tversky 1979). More precisely this means that payoffs are affected so that people view monetary gains with a decreasing value from the reference point, whereas losses are viewed as increasing. Regarding probabilities, people tended to overweigh low probabilities and undervalue high probabilities, something that is more accentuated when the probabilities pertain to a loss (Kahnemann and Tversky 1992). Central is therefore an agent's reference point. The reference point defines the so-called "frame". An agent with a reference point that makes understands its prospect more negative operates under a "losses frame", whereas an agent with a positive reference point is acting under a "gains frame". Hence, for an analysis with prospect theoretical assumptions, a player's reference point is central.

Several scholars have discussed how deterrence potentially is affected when standard game theoretical assumptions are replaced by prospect theoretical assumptions (Levy 1996, 1997; Berejikian 2002; Carlson and Dacey 2006; Butler 2007). Article I investigates what happens to a deterrence game (Chicken) when prospect theoretical assumptions are used instead of the standard concept of maximizing expected utility and what this implies for a deterrence theory.

When prospect theoretical foundations are applied to the game, it partly changes the analysis. The monotone increase/decrease changes the payoffs, but since the increase/decrease does not impact the preference order, the pure strategies remain the same as with standard assumptions. The mixed strategies do change, but only to a degree (article I, figure 3). These results in article I are similar to a study by Carlson and Dacey (2006) about the Perfect Asymmetric Deterrence game with a prospect theoretical utility function. What is different in that study is that the game is an extensive form Bayesian game and therefore several equilibria are affected, but in similar manner to the study in article I. The changes, while small, may potentially be important when studying a particular situation, but it is questionable if it does anything to change the intuitions about a theory of deterrence; article I makes the case that it does not. For a theory, it is important to be centered on what rational players would do; statistical divergence is important, but rather compliments the general picture. What is important is the introduction of psychological behavior into the research on deterrence in a

systematic way.

As mentioned, an alternative to study how people make *choices* is to study how people choose when they *play a game*. Experiments show that people are limited in reasoning about other people's behavior. The experiments reveal that such abilities differ within a group where some take strategic deliberation of the opponent into account to a larger extent than other players tend to do. The level-*k* model systematizes player behavior in levels according to their cognitive sophistication (Stahl 1993). The first level, type 0, is often defined for unpredictable player behavior where a player randomizes between possible strategies. Type 1, the next level, believes that every player is of a lower cognitive type ($k - 1$), i.e., a type 0 player, and therefore plays best reply to it. Further, type 2 plays best response to type 1, and type 3 plays best response to type 2 and so on (Article II, figure 2). Level-*k* was not constructed from experiments, like prospect theory. Rather, people playing games in certain settings seem to behave in a way that the level-*k* model captures. Over the years, several experimental studies have used the level-*k* model to explore the strategies employed by people (see for instance, Arad and Rubinstein 2012; Alaoui, Janezic and Penta 2020).

Applying the level-*k* model to a deterrence game such as Chicken changes the game. It departs from the Nash equilibrium analysis but the structure of analysis is different. Here, the first level, type 0, randomizes between the two strategies, which makes the likelihood of conflict higher than what a rational player would accept. Type 1's best reply to type 0 is therefore to cooperate. Knowing that type 1 chooses in this manner type 2's best reply is therefore to defect, and type 3's best reply to type 2 is consequently to cooperate (Article II, figure 2). Article II, uses this insight to justify Schelling's idea of irrational threats by randomizing between strategies, but in the context of this discussion, what is more important is the alternative equilibria analysis that the level-*k* model suggests (Schelling 1993).

Article I and article II make the case that regardless of representational form, the rationality assumption cannot be dispensed with when constructing a theory since this must be a part of a behavioral description. Rational behavior represents an upper bound for how an actor could behave, and a theory of deterrence must account for that. However, only accounting for rational behavior is also clearly limiting. Both prospect theory and level-*k* models offer alternative ways of capturing different behavior in a systematic way. This makes them reasonably manageable when dealing with issues such as theory and representation. Articles I and II show what reasonable areas for their respective applicability are by applying them to a classic deterrence game.

The use of statistical data from the experiments is an important method to appraise the predictions of games and examine their original assumptions. Fundamentally, their use is not unlike the study of empirical records of deterrence

situations. What differs is that while the empirical records provide an estimation of how well the model captures real world phenomena, the experiments reveal how the model's theoretical assumptions converge with how people tend to make decisions. However, how such alternative re-modelling should be appraised is debated. Replication of some of the experimental findings has been pointed out as wanting. Glenn Harrison and Don Ross (2017) show that with regards to prospect theory it is unclear how solid the data that underpin the project actually are, and advise caution (see article I for details). Such critique is of interest since it shows that just as the game theoretical model is driven by a purpose, so are the statistical models used to replace the theoretical assumptions. How the experiment is set up and which data then are used is influenced by what the goal of such remodeling actually is. This does not mean that the statistical model from an experiment is arbitrary, only that just as a theoretical model is dependent on the context in which it is put to use and its original purpose, so is the statistical model. A good example of how this matter is the application of prospect theory to deterrence. A legitimate question is whether the reweighing of the utility function to a prospect theoretical function in reality captures a more realistic behavior or simply shows that behavior potentially could diverge given a specific set of circumstances.

6 Imperfect Deterrence

In the introduction, we asked the question: what is the game theoretic model's contribution to the study of deterrence? Taking a step back and considering all the models, theories and evidence that exist to describe, relate, predict, generalize, validate and inform about deterrence it is possible to say something more general about the study of deterrence and the scientific traditions and ideals that defines it.

Clearly, the game theoretic model not only can be used, but also is used in a variety of ways throughout deterrence research. Amongst the earlier deterrence theorists, the model primarily functioned as a mediator of conditional truths. The model Chicken does not reflect the deterrence stability of the Cold War (see the exchange between Lawson 2013 and Zagare 2013). However, together with a mode like Entry Deterrence they relay key insight to what deterrence can be. The theories were formulated with reference to the models but seem to have worked primarily as inspirational reminders of important behavior traits, that could be important. Theory and the model are disconnected, which is a problem if one strives for model-theory coherence, like later deterrence theorists do. The narrow scope of game theory is also disregarded when needed to capture alternative behaviors. Thus, in this tradition, the model's contribution is largely

to serve as inspiration and backdrop for theorizing.

Relating to the game theoretic model in this way is quite distinct from how the later contributions, such as Perfect Deterrence and Nuclear Deterrence, use models. Perfect Deterrence is largely defined by the critique of earlier lack of model-theory coherence, the faulty models and the imprecise theoretical reasoning. It strives for coherence between model and theory. Perfect Deterrence satisfies many of the requirements typically associated with this ideal; the models are coherent with the theory and it solves many of the problems that have been prevailing in deterrence research. Further, it is also with this tradition that the model predictions are put to the test by statistical testing, where it seems like parts of it (Perfect Asymmetric Deterrence), accurately represents historical cases of deterrence. It has also been remodeled with prospect theoretical assumptions and while this affects the model, it does not alter the general insight it provides. Hence, at least with regards to conventional asymmetric deterrence, the theory seems relatively solid. With regards to the theory's modelling of mutual nuclear deterrence it is less clear to what extent it holds, but given that it is based on the same game-form as the Perfect Asymmetric Deterrence game, there is at least some reason to think that it may capture this sort of dynamics too.

Something similar could be said for Powell's Nuclear Deterrence, but just as any deterrence theory concerned with nuclear deterrence, it cannot be statistically validated. It sees mutual deterrence as completion in risk-taking, which is different from the Perfect Mutual Deterrence game that models credibility. Hence, as theories they suggest different dynamics of mutual deterrence. What both theories do provide are models that are coherent with their respective theories.

Hence, the contribution of a given model is largely defined by the context in which it is being used. A problem with the later contributions is that for each well-defined project, other models, theories and explanations that might offer a different perspective are excluded. For instance, while Zagare and Kilgour's main goal is to overcome the problems with classic deterrence theory and the models associated with it, they also express dissatisfaction with Powell's project.[10] In fact, there is an array of deterrence models that are in one way or another associated with the larger concept of deterrence, but do not explicitly belong to a particular deterrence theory. For instance, Catherine Langlois and Jean-Pierre Langlois (2006) propose a deterrence model that allows for bargaining. If we add to that the possible versions of any deterrence model that is altered with other types of assumptions such as those offered by prospect theory or cognitive hierarchies, the list of models without a specific place grows. What are we to make of the models

[10] Zagare and Kilgour are critical of Powell's account and label it as deterrence paradox II. However, one could argue that the Perfect Deterrence deals with general deterrence while Nuclear Deterrence is concerned with immediate deterrence, see previous section for the distinctions.

of deterrence that do not belong to any of the well defined theories?

As much as the models contribute to the various contexts, reconciling them is not possible. Alternatively, one could de-emphasize the connection to a specific deterrence theory and put the focus on the models – all models; from Chicken and Entry Deterrence (see article III) to Perfect Deterrence and Nuclear Deterrence. Philosophers call this view perspective realism (Massimi 2018, 7). In this view, we can relate to the models as a collection of tools where we may choose the ones that best suit our purposes. The relation between models is then either by *plurality*, i.e., a number of the models are in play in a given scientific context, of *partiality* – that each of the models provides a fractional account for the phenomenon, or of *complementary*, i.e., the models together map the phenomenon in a heuristic function that is defined by the scientific contexts in question (Massimi 2018, 15). This view connects to the meta-model in article III, where target criteria and purpose could be viewed as a heuristic when exploring a specific context. The problem with this view is that the idea of a deterrence theory becomes less clear. If every model is a part of one deterrence theory, then the boundaries of this theory are blurred. At worst, as argued in article II, theory is entirely lost and the only interest is the various models.

Article II makes the case that there may be a way of retaining the ideal of model-theory coherence, but letting this ideal exist alongside ideals of problem solving and development. It is the conjunction of these ideals, model-theory coherence and problem solving that has enabled the formulation of the two stronger programs of Nuclear Deterrence and Perfect Deterrence. Aiming for theory-model coherence is, as it often is with ideals, perhaps best kept that way, as an ideal. If made to be a requirement, it could potentially hogtie future attempts to solve future problems. However, if one recognizes that a situation contains attributes similar to, for instance, the Perfect Asymmetric Deterrence model, then the accompanying theory will provide a more nuanced and deep understanding of the situation, compared to, for instance, a simpler model that is not attached to a fully developed theory. In this way, the more developed approaches such as Perfect Deterrence and Nuclear Deterrence provide a deeper and fuller understanding of their respective dynamics. Simpler models will only provide a dynamic structure, where context will be missing. Hence, the heuristic view of perspective realism does suggest a theoretical approach to how we can relate to the full body of models without excluding certain types of modelling.

As argued in article II, simply adopting both model-theory coherence as well as problem-solving might however not be a tenable solution since the two may interfere with one another. When such a situation arises, one can remind oneself of what the purpose of the pursuit actually is. If the goal is to develop theory, then an ideal of model-theory coherence is reasonable to adopt. On the other hand, if

one is trying to solve a problem, then an ideal of problem-solving is reasonable to assume.

The many deterrence models with their various deficiencies and shortcomings, ingenuity and originality, the attempts to formulate deterrence theories and discussions of evidence, paints a picture of a scientific endeavor that is clearly less than perfect. However, it also bears the mark of foresight and scientific progress. Problems such as credibility have been driving forces for the development of new concepts, models and approaches. "Faulty" models such as Chicken come back in spite of their inaccuracies. Off–equilibrium behavior and bounded rationality modelling have been adopted by the deterrence community, but it is not yet clear what status such models should be given. All this reflects a tradition that actively grapples with its subject matter. This might in a sense be the game theoretic model's most important contribution to deterrence research. For all its abstraction, the model provides something concrete for the scientist to relate to, develop, expand upon or reject.

Increasingly, philosophers have come to pay attention to how scientists actually relate to the research that they are concerned with and less with what ought to be the ideals guiding them. The imperfect nature of deterrence research, its challenges, solutions and developments that have been the focus of this thesis dissertation points to the fact that this is neither strange nor wrong. They are consequences of relying on economic modelling. Catherin Elgin makes the case that inaccuracies and idealization are not signs of inadequacy. Rather, we should understand such deficiencies as promoting epistemic developments. By reviewing, discussing and reconsidering models, theories and evidence, a research tradition is forged (Elgin 2017).

This thesis introduction has tried to show the relative nature of applying game theory to deterrence research. It is clear that this is conditioned by the context and purpose of a given study. So, while the status of a model may vary throughout the research field, it is not an arbitrary type of application. It suggests a dynamic that given the specified circumstances can reveal central dynamics of deterrence interaction. Conversely, a notion of the game as the only realistic representation of deterrence would not only be naïve, it would be a view that simply is not in touch with how research of deterrence has progressed. Rather, it is a conglomerate of many different types of studies, where ideals, contexts and purposes vary. Acknowledging the imperfect nature of this type of scientific progress offers a better and more reasonable perspective. Such a view should be more than a consolation for the troubled deterrence scholar disappointed in the shortcomings of her endeavors; it is a view that represents a more sensible perspective on where deterrence research has come from and possibly, where it might go in the 21st century.

7 Article summaries

Article I: "Prospects of Deterrence: Deterrence Representation, Theory and Evidence"

Game theoretic analyses of deterrence have been criticized for not adjusting to how actors realistically behave. It has been alleged that a change to models based on prospect theoretical concepts is needed for constructing a more realistic deterrence theory. Several studies have considered prospect theory (PT) as an alternative to deterrence modelling. The article begins by analyzing a classic deterrence game, the mutual deterrence game, also known as Chicken (with standard assumptions). It then connects this game to the broader theorizing about deterrence modelling. This is done to appraise what changes occur when the standard utility function is replaced by a PT assumption. The article remodels Chicken with PT assumptions. In order to appraise the differences three scenarios, based on different reference points, are considered. The results show that with pure strategies, PT does not change the dynamics of the Mutual deterrence game, since the monotonic increase/decrease does not affect the preference order for the player. When studied with mixed strategies there is a relatively clear difference depending on what reference point a player has. The results are largely similar to the changes found when remodelling other deterrence games with PT assumptions, such as the Perfect Asymmetric Deterrence game. The differences that exist are due to the Bayesian game-form in the latter case. In neither case does a change from standard representations provide a different or new concept of deterrence.

The article then examines the experiments that underpin PT. While it seems difficult to repeat the experiments, similar estimates have been attained in comparable set-ups. Hence, simply applying PT to deterrence should be done with caution since some of the reported effects may not exist or be exaggerated. Suggesting the reconstruction of an entire theory based on such research is overstating what can be claimed. Second, other types of experiments have found that PT is outperformed by other types of boundedly rational models. Hence, it is not clear why PT should take precedence over other behavior model with empirical support. It is argued that the need to analyze a phenomenon idealized, like in classic game theory, is central for developing a deterrence theory. The contribution of PT does however enrich a theory of deterrence by opening up for psychological mechanisms that can be associated with deterrence. Consequently, rather than changing anything fundamental, prospect theory and other bounded rational model broaden the perspective on deterrence and open up for a new type of theorizing.

Article II: "A Misfit Model: Bounded Rationality and Deterrence Representation"

Contemporary theories of deterrence place a strong emphasis on model-theory coherence. As a consequence, earlier accounts of deterrence are excluded from these theories. An example is Schelling's idea of irrational threats for successful deterrence. Schelling abandons the rationality assumption embedded in game theory to explain how a player can deter by behaving irrational, but he relies on the rationality assumption to account for why actors are deterred. Such an account has been found wanting since as a solution concept it is inconsistent. However, while theoretically inconvenient, it is possible that an actor that behaves irrational, randomizing between strategies in reality would deter an opponent since such behaviour increases the possibility of conflict to a level that may be unacceptable to an opponent. The article defends and remodels Schelling's idea by employing the level-*k* model. The model is based on experiments that show that that people seem to have limited ability to reason about other people's behavior. The level-*k* model defines player behavior in stages, or levels, after their cognitive sophistication. The first level, type 0, is normally defined for simple or unpredictable player. Type 1, think that every player is a type 0 player, and in turn, plays best reply to type 0. Further, type 2 plays best response to type 1, and type 3 plays best response to type 2 and so on. The number of levels in a given analysis is dependent on the context and game in question. Since the model relaxes the standard rationality definition and could therefore give a more realistic account of a certain type of strategic behaviour. When applied to a Deterrence game such as the mutual deterrence game, the model reveals that an unsophisticated player that is randomizing over its strategy set brings about an advantageous outcome, like Schelling suggested – but without mixing solution concepts. While this means Schelling's idea can be retained, it is still problematic how we should view contributions of bounded rationality in relation to current deterrence theory, given the emphasis they place on the model-theory coherence.

There seems to be an impasse where one either has to contend with a more a coherent model-theory relationship, but exclude some accounts. Alternatively, allow for a multitude of models that better cover the phenomenon of deterrence, but undermine the connection to a theory. The article argues that the reasonable a solution is to allow for other ideals than model-theory coherence. Allowing problems and practices to also be ideals enables a larger heuristic where more advanced deterrence theories can be developed. In this version, established theories should be regarded as more developed accounts of deterrence, whereas versions that do not fit with these structures are allowed to exist as embryos to what could eventually become parts of a more nuanced theory of deterrence.

Article III: "Comparable Deterrence: Target, Criteria and Purpose"

There is a reoccurring debate regarding which deterrence model best represents deterrence. Comparing models is not something that the philosophic or scientific literature has previously engaged with. Rather, the focus has been on representational qualities of a model and ideals guiding model construction. Yet, the arguments for the advantages with one model over another are reoccurring.[10] As discussed in the introduction, there are various aspects of relevance when appraising a model. The article examines three aspects; the target of a model, the criteria used in the construction of the model, and the type of purpose for developing the model. The three models compared in the article are Chicken, Entry Deterrence and Perfect Mutual Deterrence. With the use of the three aspects, it is evident that the three deterrence models address different representations (target), in different ways (criteria) and to different ends (type of purpose). The aspects show that that model comparison is context dependent and therefore models are often complimentary rather than rivaling. It is concluded that much of the discussion regarding which deterrence model is superior is misguided. For example, some of the detractors of later model developments take the view that the Perfect Mutual Deterrence model provides nothing new, that other models cover the same target, and that they do it better. This line of argument can be shown to be invalid. The Perfect Mutual Deterrence model should rather be recognised as a clear advancement in the research field. However, while the Perfect Mutual Deterrence model is a clear development, it should not, and cannot automatically replace other deterrence models, as suggested by some. Rather, a plurality of models should be viewed favourably within the research community.

The article does essentially two things: it suggests a framework for model-comparison and it settles a debate. To settle a debate is of course philosophically satisfying, however, more important is the framework developed in the article. With the framework, model-comparison can be made more intelligible. By analyzing a model by its target, criteria and type of purpose models can be better compared and analyzed.

[10] This debate primarily consists of Stephen Walt's 1999 article "Rigor and Rigor Mortis – Rational Choice and Security Studies" and Frank Zagare's reply in the same year "All Rigour No Mortis", Stephen Quackenbush's discussion from 2011a entitled "Deterrence Theory – Where Do We Stand?", Fred Lawson's critique from 2013 of Quackenbush's article entitled "Back to the Future in the Study of Deterrence", and Frank Zagare's reply to Lawson in 2013 "Deterrence Then and Now - There Is No Going Back".

Article IV: "Deterring the Dauntless: Appraising the effects of naval deterrence against the Somali piracy"

The article studies a situation where deterrence was the defined goal of a military operation. During the period 2009–2013, several naval operations attempted to deter the piracy emanating from Somalia. The article uses the so-called Perfect Asymmetric Deterrence game, in order to analyse the deterrence that the navies tried to achieve. Since the model proposes four different outcomes, the level of engagement between pirates and navies is calculated. The piracy is measured by incident frequency. The naval counter-piracy activities are measured by the number of disrupted pirate vessels. The outcomes are then examined in two statistical models for the two areas of operation (Gulf of Aden and Somali Basin): one without control variables and one which controls for the shipping community's own security precautions, and the security political situation in Somalia. Because there are often several outcomes associated with each equilibrium, the equilibria are analyzed and interpreted. It is found that the naval operations did deter the Somali pirates in the Gulf of Aden, but that this objective was not met in the Somali Basin. It is concluded that the navies' operational focus on the Gulf of Aden, coupled with the fact that that area is relatively smaller than the Somali Basin, made the naval threat credible. The credibility of the threat effectively denied the pirates from freely operating in the Gulf of Aden. Conversely, limited attention by the naval units and the long Somali southern coast with open waters impeded naval control in the Somali Basin.

The deterrence model used captures possible outcomes between two agents where one tries to deter the other. Although originally not developed to study deterrence on a operational level, the model does capture the strategic interaction between pirates and navies well. However, the study required a lot of data in order to become robust. Hence, data accessibility is potentially a disabler for future similar studies. Since there are several assumptions embedded in the study, one must also caution not to over-interpret the results. Nevertheless, the fact that the game model used could be applied to a different situation than it was developed for is interesting as it indicates that it does capture something fundamental in deterrence interaction.

8 References

Alaoui, Larbi, Janezic, Katharina A. and Penta, Antonio. 2020. "Reasoning about others' reasoning", *Journal of Economic Theory*, Vol. 189: 1–51.

Arad, Ayala and Rubinstein, Ariel. 2012. "The 11-20 Money Request Game: A Level-*k* Reasoning Study", *American Economic Review*, Vol. 102, No. 7:3561–73.

Brodie, Bernhard. 1959. *Strategy in the Missile Age*. Princeton University Press.

Camerer, Colin. F. 2003. *Behavioural Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Carlson, Lisa and Dacey, Raymond. 2006. "Sequential Analysis of Deterrence Games with a Declining Status Quo", *Conflict Management and Peace Science*, Vol. 23, No. 2:181–198.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford University Press.

da Costa, Newton C. A. and French, Steven. 1990. "The Model-Theoretic Approach in Science", *Philosophy of Science*, Vol. 57, No. 2: 248–265.

Creath, Richard. 1987. "The Initial Reception of Carnap's Doctrine of Analyticity", *Noûs*, Vol. 21, No. 4:477–499.

Elgin, Catherine, Z. 2017. *True Enough*. MIT Press.

Field, Alexander J. 2014. "Schelling, von Neumann, and the Event that Didn't Occur", *Games*, Vol. 5, No. 1:53–89.

van Fraassen, Bas. 1989. *Laws and Symmetry*. Oxford University Press.

Freedman, Lawrence. 2004. *Deterrence*. Polity.

George, Alexander and Smoke, Richard. 1974. *Deterrence in American Foreign Policy: Theory and Practice*. Columbia University Press.

Giere, Ronald. N. "Using Models to Represent Reality" in *Model-Based Reasoning in Scientific Discovery*. Magnani, L., Nersessian, N.J. and Thagard, P. (eds). 1999. Springer, pp. 41–57.

Giere, Ronald. N. 2004. "How Models Are Used to Represent Reality", *Philosophy of Science*, Vol. 71, No. 5:742–752.

Grüne-Yanoff, Till and Schweinzer, Paul. 2008 "The Role of Stories in Applying Game Theory", *Journal of Economic Theory*, Vol. 15, No. 2:131-146.

Grüne-Yanoff, Till. 2013. "Appraising Models Nonrepresentationally", *Philosophy of Science*, Vol. 80, No. 5:850–861.

Harrison, Glenn W. and Ross, Don. 2017. "The Empirical Adequacy of Cumulative Prospect theory and Its Implications for Normative Assessment", *Journal of Economic Methodology*, Vol. 24, No. 2:150-165

Hesse, Mary. 1966. *Models and Analogies in Science*. University of Notre Dame Press.

Homer. 2009. *The Iliad* (translated by K. S. Klein). Penguin.

Huth, Paul and Russett, Bruce. 1984. "What makes deterrence Work? Cases from 1900 to 1980", *World Politics*, Vol. 36, No 4:496–526.

Huth, Paul and Russett, Bruce. 1988. "Deterrence Failure and Crisis Escalation", *International Studies Quarterly*, Vol. 32, No. 1:29–45.

Huth, Paul. 1988. *Extended Deterrence and the Prevention of War*. Yale University Press.

Huth, Paul. 1999. "Deterrence and International Conflict: Empirical Findings and Theoretical debates", *Annual Review of Political Science*, Vol. 2, 61–84.

Intriligator, Michael D. and Brito, Dagobert L. 1984. "Can Arms Races Lead to War?" *Journal of Conflict Resolution* Vol, 28, No. 1:63–84.

Isaac, Joel. 2005. "W. V. Quine and the origins of Analytic Philosophy in the United States", *Modern Intellectual History*, Vol. 2, No. 2:205–234.

Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton University Press.

Kahn, Herman. 1960. *On Thermonuclear War*. Princeton University Press, reprinted by Transaction Publishers 2007.

Kahneman, Daniel and Tversky, Amos. 1979. "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, Vol. 47, No. 2: 263–291.

Kahneman, Daniel and Tversky, Amos. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty", *Journal of Risk and Uncertainty*, Vol. 5. No. 4:297–323.

Kaplan, Fred M. 1983. *The Wizards of Armageddon*. Stanford University Press.

Keeny, Spurgeon M. Jr. and Panofsky, Wolfgang, K. H. 1981. "MAD versus NUTS: Can Doctrine or Weaponry Remedy the Mutual Hostage Relationship of the Super Powers", *Foreign Affairs*, Vol. 60, No. 2: 287–304.

Langlois, Catherine and Langlois Jean-Pierre. 2006. "Bargaining and the Failure of Asymmetric Deterrence: Trading off the Risk of War for the Promise of a Better Deal", *Conflict Management and Peace Science*, Vol. 23:159–180.

Lawson, Fred. 2013. "Back to the Future in the Study of Deterrence", *St Anthony's International Review*, Vol. 9 No. 1:144–156.

Lebow, Richard and Gross Stein, Janice. 1990a. *When Does Deterrence Succeed and How Do We Know?* Canadian Institute for International Peace and Security.

Lebow, Richard and Gross Stein, Janice. 1990b. "Deterrence: The Elusive Dependent Variable", *World Politics*, Vol. 42, No 3:336–369.

Lieberman, Elli. 1994. "The Rational Deterrence Theory Debate: Is the Dependent Variable Elusive?" *Security Studies*, Vol. 3, No 3: 384–427.

Luce, Duncan R. and Raiffa, Howard. 1957. *Games and Decisions: Introduction and Critical Survey*. Reprint 1989. Dover Publications.

Mäki, Uskali. 2009. "MISSing the World – Models as Isolation and Credible

Surrogate Systems", *Erkenntnis*, Vol. 70, No. 1: 29–43.
Massimi, Michaela. 2018. "Perspectival Modelling" *Philosophy of Science*, Vol. 85, No. 3:335–359.
Morgan, Patrick M. 1983. *Deterrence: A Conceptual Analysis*. 2nd edition. Sage.
Morgan, Patrick. 2003. *Deterrence Now*. Cambridge University Press.
Morgan, Mary S. and Knuuttila, Tarja. 2008. "Models and Modelling in Economics", in *Handbook of the Philosophy of Economics*. Mäki, Uskali (ed). *Forthcoming*
Myerson, Roger B. 1991. *Game Theory: Analysis of Conflict*. Harvard University Press.
Nalebuff, Barry. 1986. "Brinkmanship and Nuclear Deterrence: The Neutrality of Escalation" *Conflict Management and Peace Science*, Vol. 9, No. 3:19–30.
Narang, Vipin. "What Does it Take to Deter? Regional Nuclear Postures and International Conflict", *Journal of Conflict Resolution*, Vol. 57, No. 3:478-508.
Powell, Robert. 1990. *Nuclear Deterrence: The Search for Credibility*. Cambridge University Press.
Quackenbush, Stephen, L. 2010. "General Deterrence and International Conflict: Testing Perfect Deterrence Theory", *International interactions*, Vol. 36, No. 1: 60–85.
Quackenbush, Stephen, L. 2011a. "Deterrence Theory: Where Do We Stand?" *Review of International Studies*, Vol. 37, No. 2: 741–762.
Quackenbush, Stephen, L. 2011b. *Understanding General Deterrence: Theory and Application*. Palgrave MacMillan.
Quackenbush, Stephen and Zagare, Frank, 2016. "Modern Deterrence Theory: Research Trends, Policy Debates and Methodological Controversies", *Oxford Handbooks Online*. Oxford University Press.
Rauchhaus, Robert. "Evaluating the Nuclear Peace Hypothesis: A Quantitative Approach", *Journal of Conflict Resolution*, Vol. 53, No. 2:258-277.
Rubinstein, Ariel. 2006. "Dilemmas of an Economics Theorist", *Econometrica*, Vol. 74, No. 4: 865–883.
Russell, Bertrand. 1959. *Common Sense and Nuclear Warfare*. Simon and Schuster.
Schelling, Thomas. C. 1960. *The Strategy of Conflict*. Harvard University Press.
Schelling, Thomas. C. 1966. *Arms and Influence*. Yale University Press.
Selten, Reinhard. 1978. "The Chain Store Paradox", *Theory and Decision*, Vol. 9, No. 2: 127–159.
Slantchev, Branislav L. 2012. *Military Threats: The Costs of Coercion and the Price of Peace*. Cambridge University Press.
Snyder, Glenn. 1961. *Deterrence and Defense*. Princeton University Press.

Snyder, Jack. 1976. "Rationality on the Brink - The Role of Cognitive Processes in Failures of Deterrence", *RAND Paper Series*.

Sörenson, Karl. 2019. *In Search of Lost Deterrence: Two essays on deterrence and the models employed to study the phenomenon*. Licentiate thesis, KTH Royal Institute of Technology.

Stahl, Dale O. 1993. "Evolution of Smart n players", *Games and Economic Behavior*, Vol. 5:604–617.

Suárez, Mauricio and Cartwright, Nancy. 2008. "Theories: Tools versus Models", *Studies in History and Philosophy of Modern Physics*, Vol. 39, No. 1:62–81.

Sugden, Robert. 2000. "Credible Worlds: The Status of Theoretical Models in Economics", *Journal of Economic Methodology*, Vol. 77, No. 1:1–31.

Suppes, Patrick. 1960. "Models of Data" in *Logic Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*, Nagel, Earnest, Suppes, Patrick and Tarski, Alfred (eds.) Stanford University Press, pp. 252–261.

Vegetius Renatus, Pubilus Flavius. 2012. *De Re Militari* (translated by John Clarke). Leonaur.

Walt, Stephen. 1999. "Rigor or Rigor Mortis? Rational Choice and Security Studies", *International Security*, Vol 23, No. 4:5–48.

Weisberg, Michael. 2013. *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press.

Wirtz, James J. 2018. "How Does Nuclear Deterrence Differ from Conventional Deterrence?" *Strategic Studies Quarterly*, Vol. 12, No 4:58–75.

Woosang, Kim and Bueno de Mosquita, Bruce. 1995. "How Perceptions Influence the Risk of War", *International Studies Quarterly*, Vol. 39, No. 1:51–65.

Zagare, Frank and Kilgour, Mark. 1993. "Asymmetric Deterrence", *International Studies Quarterly*, Vol. 37, No. 1: 1-27.

Zagare, Frank. 1999. "All Rigor, No Mortis", *International Security*, Vol 24, No 2:107-114.

Zagare, Frank. 2013. "Deterrence Theory Then and Now: There is No Going Back", *St Anthony's International Review*, Vol. 9. No. 1:157-167.

Zagare, Frank and Kilgour, Mark. 2000. *Perfect Deterrence*. Cambridge University Press.