In Search of Lost Deterrence

- Two essays on deterrence and the models employed to study the phenomenon

Karl Sörenson*

2018

Licentiate Thesis in Philosophy
KTH Royal Institute of Technology
Stockholm, Sweden

* To contact the author, please refer to karl.sorenson@fhs.se or Karl Sörenson, Drottning Kristinas väg 37, PO-BOX 27805, 11593 Stockholm

Contents

Abstract

To deter is central for strategic thinking. Some of the more astute observations regarding the dynamics of deterrence were made during the Cold War by game theorists. This set the stage for how deterrence has come to be studied. A strong methodological element like the research on deterrence's reliance on game theory requires examination in order to understand what sort of knowledge it actually yields. What sort of knowledge does one acquire when deterrence is viewed through game theoretic models? How do they inform us about the phenomenon of deterrence? To understand the nature of a phenomenon through models requires idealization, which in turn presupposes assumptions. This licentiate thesis investigates the type of knowledge we attain when approaching deterrence from a game theoretic perspective. The two articles presented address two separate but related issues. The first article reviews a debate regarding which deterrence model best capture the phenomena of deterrence, i.e. how models can be compared to one and other. The article presents a framework for comparing models and then appraises how these different deterrence models inform us about deterrence. The second article uses one of the more central deterrence models in order to evaluate how and to what extent the naval operation Atalanta managed to deter the Somali piracy.

Svensk sammanfattning

Avskräckning är centralt för strategiskt tänkande. Några av de mest skarpsinniga observationerna angående avskräckning gjordes av spelteoretiker under Kalla kriget. Denna typ av forskning har kommit att definiera hur avskräckning studeras. Ett stark metodologisk angreppssätt så som spelteori, kräver granskning för att förstå vilken typ av kunskap som faktiskt erhålls om ett komplext fenomen som avskräckning. Vilken typ av kunskap får vi när avskräckning förstås ur ett spelteoretiskt perspektiv? Hur gör det oss klokare om avskräckning som fenomen? Att förstå ett fenomen genom de modeller vi använder oss av kräver någon typ av idealisering, vilken i sin tur bygger på antaganden. Denna licentiatavhandling undersöker vilken typ av kunskap vi får när man närmar oss ett fenomen som avskräckning genom spelteori. De två artiklarna som presenteras redogör för två separata men relaterade frågor. Den första artikeln presenterar ett ramverk för att jämföra avskräckningsmodeller varpå dessa analyseras för att förstå vilken sorts kunskap de olika modellerna förmedlar. Den andra artikeln använder en av de mer centrala avskräckningsmodellerna för att utvärdera hur och i vilken utsträckning den marina operationen Atalanta lyckades avskräcka de somaliska piraterna.

Acknowledgements

L'homme n'est qu'un roseau, le plus faible
de la nature, mais c'est un roseau pensant.[1]
Blaise Pascal, *Pensées*, 1662

1. Deterrence and Scientific Models

A firm lowers its prices to impede competition. Law enforcement intensifies patrolling in an area to dissuade crime. A country increases its military expenditure to signal commitment to its defenses. The goal to deter is important to society; its businesses, its institutions – all handle elements of deterrence. For a defense force, it could be argued that deterrence is the main activity. The concept of deterrence is old. The Roman dictum *si vis pacem, para bellum*[2], indicates that it is ancient. Yet, the fundamental will to protect what we value and dissuade activities that might be harmful suggests that deterrence as a concept is not merely ancient, it is a behavioral strategy. Its importance to humanity has made deterrence a well-studied subject. Many disciplines engage with it, philosophy, economics, political science, psychology, criminology, biology – after all, deterrence is not mere a human activity.

To study deterrence through the decisions it entails in the form of a game has been a recognized scientific approach since the 1950's. During this period, game theory, nuclear strategy and research on atomic weapons developed in tandem and the subjects seemed to have drawn upon one another for inspiration and development (see, for instance, Meyerson 1991). A common approach across the disciplines that study deterrence, but perhaps most notably in economics and political science, is to identifying choices, actors and preferences, which define the dynamics in a deterrence situation. This means that in its most basic form, an actor deters when it makes another actor refrain from performing a certain action. More specifically: "*A* refrains from *x,* because *B* threatens to do *y"* (where *A* and *B* are decision-making agents, *x* and *y* are actions). Because deterrence is about preventing a certain action *x*, by threatening to perform a certain action *y*, if *A* carries out action *x* in spite of *B* threatening to do *y*, then deterrence has failed. Failure of deterrence is as much part of the studied phenomenon as its success. Further, it is of equal interest to assess, in case action *x* was not carried out by *A,* whether this is due to *B's* threat of doing *y*, or because of some other reason. If the latter is the case, then there is no deterrence in play, but these types of situations must also be part of the analysis, albeit separated from the deterrence. Hence, already at this very basic level, deterrence is complex to study. However, as we will see, this is only one part of what constitutes deterrence that the scientific community is engaged with. It is deterrence and the models used to study it that is the main topic of this thesis.

---

[1] "The human being is only a reed, the most feeble in nature; but it is a thinking reed" (Author's translation).
[2] "If you want peace, prepare for war".

Given the initial form "*A refrains from x, because B threatens to do y"*, it should be noted that deterrence has predominantly been studied as a game theoretic problem rather than as a decision theoretic problem. There is an important reason for this. A decision about defending against a possible attack or not defending will need to entail ideas about how an opponent might reason. For instance, an agent pondering an attack on another agent must take into account how that agent will respond. If the agent considering the attack believes that the other agent will defend itself, this will entail one outcome; if it believes that the other agent will concede to an attack that will entail another outcome. Which outcome will be the result of this interaction is in part dependent upon what the first agent decides to do, and in part, on what the other agent decides to do. Similarly, the agent suspicious of an attack will recognize that how it decides to react to an attack will be an important part of the decision that the aggressor is contemplating. Consequently, decisions made by one agent will influence the decisions of the other, and vice versa, and the outcome will be the result of both agents' choices. The deliberations will in part be replies to what is believed to be the choice of the other. Hence, studying deterrence as a decision against nature will lose the crucial aspect of strategic interaction that goes to the heart of deterrence.[3]

To view deterrence in terms of game theoretic models makes sense since this will take all parties' intentions and preferences into account. This is essential if one wants to understand the strategic nature of deterrence. Yet there is a significant distance between the movements of troops, the positioning of nuclear strike submarines and the augmentation of defense budgets to the games we employ to represent deterrence. When approaching deterrence through game theory there is a need to grasp what is left in the model representing the world, and what we leave out. What sort of knowledge does one acquire when deterrence is viewed through the prism of game theory? How does it inform us about the phenomenon of deterrence? These are philosophical questions. To understand the nature of a phenomenon through models requires some sort of idealization, which in turn presupposes some underlying assumptions. To examine such assumptions and practices is a philosophical enquiry into the foundation of scientific knowledge. A strong methodological element like the research on deterrence's reliance on game theory warrants examination in order to explore what sort of knowledge it yields.

From a practitioner's perspective, the issue of understanding the scopes and limits of game theoretic analysis is pressing since game theory is put to use in the analysis of national defense as well as in the design of the security systems that defines it. What the research on deterrence ultimately aims to do is to inform policy making about potentially dangerous situations

---

[3] See for instance Richard Jeffrey (1983) who discusses deterrence as a decision.

from crisis-bargaining to nuclear war. Thus, the sheer magnitudes involved motivate investigation. The aim of this thesis is to contribute to the question of deterrence, by investigating the models we employ to study it. On the one hand, this enables distinctions of important concepts as well as refuting false claims of accuracy or overconfidence in model practices; on the other hand, what remains will provide instrumental guidance when navigating the choppy waters of security political decision making.

The licentiate thesis consists of three parts: an introduction and two articles (I and II). The first section of the introduction discusses central philosophical themes regarding scientific models. In section 2–5, important aspects of model developments regarding deterrence are appraised from the point of what sort of scientific explanation they yield. The sixth section presents conclusions.

The first article (I) presented in this thesis reviews a debate regarding which deterrence model best captures the phenomenon of deterrence. Since this is a question of comparability, i.e. how models can be compared to one another, the article presents a framework for comparing models and then appraises how these different deterrence models relate to the phenomenon of deterrence. That framework is the main contribution of the first article. The second article (II) applies one of the more central deterrence models in order to evaluate how and to what extent the naval operation Atalanta managed to deter the Somali piracy. In so doing, it shows how a deterrence model can help to make sense of a complicated violent social situation. Thus, while one article addresses model comparability and the other investigates a model's applicability, the thesis introduction addresses the larger issue of how we can understand deterrence models epistemologically.

2. What is a model?

In the study of deterrence, game theoretical models play a prominent role. But what is a model? How should we relate to the scientific knowledge that a model yields? One of the earliest accounts of economic models was put forth by John Stuart Mill. Mill did not use the term "model" himself, but his understanding of it corresponds broadly to how economists use models today (Morgan and Knuuttila 2008). The central idea to Mill was that theoretical development in economics would benefit from developing what he called "figures of thought", which theory development could employ for abstract deduction (Mill 1843). The rudimentary idea of the model is to identify causally relevant features of the phenomenon one wants to study. When representing the world in this manner many parts of the studied phenomenon are left out; it is only the parts of perceived relevance that are modelled. Scientifically this has proved a much

sought after advantage since this type of abstraction can discard unimportant aspects and focuses attention. Since a model of a given phenomenon is partial, much attention has been devoted to the representational qualities of a model. Ronald Giere argues the point of representation by highlighting the importance of a model's similarity to its target (Giere 2004). However, as a model's overarching purpose is to help the scientist in the study of a phenomenon, only the relevant properties need to go into the representation. Thus, a total of similarities between a model and its target are not the concern, it is the similarity between a model and the relevant properties which are of importance. The representation is guided by the modeller's interest (Giere 2004).[4] Representing deterrence is not different from representing any other social phenomenon. Certain aspects will be left out whereas others will be given specific focus. The interest relative aspect of a model plays a central role when considering a phenomenon. A model that aims to *explain* deterrence may need to represent the phenomenon differently from a model that is made to *predict* deterrence behaviour.

When using a scientific model we rely on a type of analogical reasoning in order to make inferences. The analogical aspect of the relationship between model and target is given specific attention by Mary Hesse who distinguished several important aspects with the analogy. Hesse focused on two aspects of the analogical reasoning; the relationship that exists *within* the object of reference (the source) and the object under investigation (the target) respectively, and the relationship *between* the source and the target (Hesse 1966, 59). By investigating the similarities and dissimilarities between the source and the target, the strength of the analogical relationship can be mapped out. The most important relationship is what Hesse calls the "neutral" relationship (Hesse 1966). This is when it is not known if the relationship is of a positive or negative nature (i.e. similar/dissimilar). Take, for instance, analogical relationship between a canary and a human. There are plenty of dissimilarities, type of species (bird/human), anatomy, body mass, metabolism, etc. In contrast, an essential similarity is that both species have similar breathing systems with lungs that provide oxygen to the bloodstream. Supposing that one wants to reach conclusions about when the berthing environment in a mine becomes dangerous, the strength of the analogy will rest on two things. First, the validity of the analogy will depend upon the positive parts of the analogy compared to the negative parts. For example, while it is true that the breathings systems are similar in their design, the size of them differs to such a significant extent that the two systems may very well behave differently. The second part of the analogical inference is the relationship between the new property and the properties already known to be positive or negative. Hesse points out that if the similarities of the analogy are dependent upon

---

[4] Uskali Mäki argues a related point to and single out *resemblance* as a central aspect of a model, i.e. that a model shares a partial resemblance to its target (Mäki 2007).

one and other in a favourable sense, the argument will be strong (Hesse 1966, 58). Thus, the new property, in this particular case, heightened levels of carbon monoxide, exposes a neutral analogical relationship. That is, the similarities and the dissimilarities are not decisive as there is a level of uncertainty about what part of the analogical relationship that prevails. Today the negative effects of heightened levels of carbon monoxide are well-established empirical facts. However, at the time when canaries were used to alert miners the inferences drawn from canary to human relied upon the analogical relationship.[5]

It is worth noting that the interest-relative aspect plays an import role in this context with regards to the new property. The quicker reaction time of the canary makes it sufficient as a predictor for humans concerning heightened carbon monoxide levels. In other contexts, the canary may be less ideal to draw inferences from. For Hesse as well as for Giere and Mäki, when drawing conclusions about the world based on inferences from a model the strength hinges largely on the similarities, dissimilarities, how they relate to the neutral relationships and what the purpose of the analogical inferences is. Similarities and dissimilarities, and especially the neutral relationship between models and the world they aim to represent will be important aspect when appraising deterrence models.

While representational qualities, such as similarity, as well as the purpose a model is made for, are central when appraising a model, an overarching question is what scientific value can be ascribed to a given model? In an influential article, Robert Sugden suggests that what is central for a model is not that it represents an actual phenomenon in a certain way; it is that it presents a plausible, or convincing, world, which informs us about some type of situation.[6] The construct of the model often accounts for how something specific functions, but with the model we learn something much more general, which goes beyond the specific target of the model (Sugden 2000, 5). From Sugden's point of view, there is a whole range of situations, which a model, in spite of limited actual target, potentially can account for. Thus, Sugden downplays actual representation in terms of similarity or resemblance as a necessary criterion; it is a model's potential to make sense of a situation, which is of interest.

Ariel Rubinstein advances a similar view to Sugden's. Rubinstein argues that aiming for actual representation is one (of several) dilemma(s) a modeller must address. Actual representation is to Rubinstein not something a modeller must strive for since other aspects can be of greater interest to investigate (Rubinstein 2006). In a similar vein, Michael Weisberg

---

[5] Hesse uses the example of whether the moon would be habitable for human beings, based on analogical inferences from the habitability of the earth. In 1966, when Hesse published her book on analogical reasoning the issue was not furnished with the same amount of data as it is today. The relationship between analogies and empirical data is similar to how models can relate to data; more facts can increase or decrease our confidence in a particular model.

[6] Sugden uses the world "credible" for describing a convincing model, however, as "credible" and "credibility" is central in the discussion of deterrence, I will use the world "plausible" or "convincing" for matters of clarity.

suggests that the process of modelling is guided by different ideals. Sometimes a model attempts to represent an actual target, sometimes its potential to represent is the focus (Weisberg 2013). An ideal can shift from making a model represent its target as simply as possible, to an ideal of incorporating several complexities to understand a phenomenon.

Till Grüne-Yanoff makes a point related to that of Sugden, but differentiates more clearly between different types of knowledge. Grüne-Yanoff sees the distinction between "how-possibly explanations" and "how-actually explanations" as a potential dichotomy for relating to how we can understand what scientific value a model has. While how-actually explanations aim to identify true causes related to a phenomenon, which explicate the object one wants to explain (the explanandum), how-possibly explanations singles out causes that possibly explain the explanandum, but not necessarily point out an actual empirically validated cause or causes (Grüne-Yanoff 2013). In this context, there is a discussion regarding how these distinctions relate to one another. In one view, how-possibly explanations and how-actually explanations are different stages of an epistemic process. From this perspective, how-possibly explanations have lower empirical support than how-actually explanations, but function as a first step. Once sufficient empirical support for a given cause has been attained (and possibly others removed) it becomes a how-actually explanation (see Resnik 1991, 41). A different view is that how-possibly explanations are distinct from how-actually explanations in that they explore possible causes for a given phenomenon, but also rule out certain causes (Dray 1957). How-actually explanations account for how a given event actually happens, pointing out the true cause.

Giere, Mäki, Rubinstein, Sugden and Weisberg never actively consider the distinction between how-possibly and how-actually explanations. Nevertheless, it seems that there is a discernible line between those who appear to think of how-possibly explanations as an equally valuable aspect of a model as a how-actually explanation, and those who tend to think of how-possibly explanations as inferior to how-actually explanation. Sugden's idea of credible worlds, Rubinstein's discussion of dilemmas, and Weisberg's discussion of guiding ideals all pertain to this idea that how-possibly explanations are of equal interest to how-actually explanations (Sugden 2000; Rubinstein 2006; and Weisberg 2013). Giere, Hesse and Mäki, on the other hand, are less distinct, but tend to underscore that representational aspects of models are important and hence focus their discussion on representation. This indicates that they may find how-possibly explanations as too hypothetical and detached from the real target. Actual knowledge of true causes for a phenomenon is largely what the scientific enterprise is about. Notwithstanding, the importance of true causes, there is a case to be made that the hypothetical knowledge would be difficult to be without, and not just as a precursor.

The extent to which a model actually represents a phenomenon, whether the model resembles its target and whether the explanations provided by the model tend to be how-possibly explanations, are important aspects when appraising what we learn from deterrence models and what sort of knowledge they provide. So what claims do they make? For the purpose of this discussion, representation, type of purpose, whether there is a analogical relationship and the sort of explanation a model provides will be central components when considering the scientific value of deterrence models.

3. Potential Deterrence

The philosopher Bertrand Russell is credited with being the first to identify the most known deterrence model commonly referred to as the *Game of Chicken,* or simply, *Chicken.* Russell suggested the idea of two hooligans driving against one another where the first one to swerve loses, as an image of the danger with a nuclear arms race between the Western Northern Atlantic Treaty Organization (NATO) and the Eastern Warsaw-pact (Russell 1959).

> "Since the nuclear stalemate became apparent, the Governments of East and West have adopted the policy which Mr. Dulles calls "brinkmanship." This is a policy adapted from a sport which, I am told, is practiced by some youthful degenerates. This sport is called "Chicken!" It is played by choosing a long straight road with a white line down the middle and starting two very fast cars towards each other from opposite ends. Each car is expected to keep the wheels of one side on the white line. As they approach each other, mutual destruction becomes more and more imminent."

Russell sketched what he perceived to be the dilemma facing the bi-polar world's nuclear competition. Russell's portrayal takes the form of a game; if none of the participants swerves, they both will crash, the analogy for nuclear annihilation. If one swerves, the one who remains will gain social status among his peers, analogous to a strategic advantage. Finally, if both swerve, they will both be fine, but none will improve on its situation (the game formally discussed in article I). While Russell was less explicit on the alternative outcomes, when one or both parties swerve, the anecdote, or game, was invoked to show the dangers with a nuclear arms race. Other scholars, some perhaps with a less critical agenda than Russell, stuck with the Chicken-portrayal as fitting for capturing the overarching strategic situation of the Cold War.

Russell's game was soon followed and further developed by two other RAND researchers, Herman Khan and Tomas Schelling.[7] Khan's tome *On Thermonuclear War* was published in 1960 and Schelling's *The Strategy of Conflict* in the same year (Khan 1960; Schelling 1960). While RAND researchers continued to focus on deterrence as a nuclear strategy by employing games and systems analysis, other scientists connected deterrence through the study of game theory to other related areas such as international relations, political science and war studies. Glenn Snyder and Paul Diesing connected deterrence to international relations (Snyder and Diesing 1979), while authors such as Robert Jervis (1976) and Jack Snyder (1976) departed from formal models such as Chicken to discuss specific psychological aspects of deterrence, derived from the model. These are just some of the more central and notable contributions as many more were to follow.

The game theoretic game *Chicken,* which became a corner stone in the strategic analysis by RAND researchers and social scientists, differs in important aspects from how Russell originally described it. Chicken, as understood by the game theoretic community, depicts a situation where two players have the same incentives to swerve, remain and make their moves simultaneously. We say that the game is symmetric. When relying on Chicken to understand the Cold War deterrence between NATO and the Warsaw-pact, some issues arise. First, Chicken deviates from the more basic form of deterrence as stipulated in the introduction, "*A* refrains from *x,* because *B* threatens to do *y*". In Chicken, both agents try to deter one another. Further, non-cooperate games, such as Chicken, where decision-making agents have conflicting interests are analyzed with the concept of Nash equilibrium. A pure Nash equilibrium is when each player's chosen strategy is the best strategy it can choose (i.e. it cannot benefit by switching to another strategy) given the other player's strategy. There are two pure Nash equilibria in Chicken; one where one player swerves and the other stays, and its complement where the other player swerves and the first stays. This is because if one player believes that the other player will stay the course; its best reply to that strategy is to swerve. If, on the other hand, the player believes that the other player will swerve, its best reply to that strategy is to stay on course. Because the game is symmetric the same reasoning applies to the other player's choice of strategy. The two pure Nash equilibria are not very informative in the Chicken game. They do rule out the possibility of both swerving (Status quo) and both staying the course (Crash), however they only inform us of that either one of the two players attains an advantage. Thus, Chicken's pure Nash Equilibria are not very revealing for predicting player behavior. There is also a mixed Nash equilibrium, where

---

[7] The same year as Russell presented his game, Bernard Brodie a strategist at the Research and Development Cooperation (RAND), formulated what a nuclear deterrence strategy implied in his *Strategy in the Nuclear Age* (Brodie 1959). In his book, Brodie outlined his thoughts on deterrence.

the Status quo (i.e. where both swerve) is likely to occur, but open up for other outcomes.[8] A mixed Nash equilibrium is when at least one of the players randomizes between the possible strategies it can take.[9] Hence, while the pure equilibria inform us that one player will always attain an advantage, the mixed equilibrium reveals a positive probability that both players will swerve, but it does not rule out that the players will clash. The dilemma of Chicken captures a crucial aspect with deterrence, namely that attempting an advantage invites a risk of all-out war, while abiding may invite the opponent to take advantage.

Given how the game theoretic Chicken portrays the strategic situation of the Cold War it differs in several crucial ways from the situation suggested by Russell, which allegedly accounts for the same situation. Thus, there are two stories, Russell's analogy and the game of Chicken that both possibly explain the strategic situation during the Cold War. Reverting to Hesse's tabulation of analogies it is possible to represent the relationship between Russell's model (the source) and the game theoretic model (Chicken) with reality, nuclear deterrence:

Russell's Model	The Chicken Model	Target
- Two speeding youngsters - Four possible outcomes - Definite order of preferences	- Two rational decision making agents - Four possible outcomes - Definite order of preferences	- Two political blocks - Several possible outcomes - Some order of preferences
- High likelihood of disastrous outcome ↓ Dangerous Instable Game	- One player gets an advantage / both remain in the Status quo with a relatively high probability ↓ Not Dangerous Instable Game	- Possibility of outcome which may lead to disaster ↓ Dangerous Stable Situation

Figure 1. Inference by analogy.

---

[8] That both players swerve is equated with Status quo is in a sense strange, because in a game of Chicken there is not really a situation where nothing is happening. Rather, all the players are in movement, and the question is if they will swerve. However, when both players swerve this will be analogous to the idea that neither of the contestants try to get an advantage.

[9] Robert Aumann develops the idea of a correlated equilibrium as an alternative solution for the Game of Chicken. The correlated equilibrium builds on the idea that the players receive a signal from a third party about which strategy they should choose (Aumann 1974).

When represented in this manner it is clear that certain aspects of the two depictions (Russell's model and the Chicken model) are similar to their target, i.e. the period of the Cold War, others are dissimilar, and some are neutral. Before turning to the target, one clarification is needed: in game theory, "stability" is a technical term with a precise meaning. A mixed Nash equilibrium is said to be stable when a small change in the probability of an event does not change the player's strategy. Hence, a game like Chicken is a stable game, when studied with mixed equilibria. However, this is not how the term is used in the discussion about deterrence in the Cold War. In this context, "stability" is used to infer that nothing is happening, that the situation is stable, i.e. Status quo remains. Thus, in this context the term "stability" applies to situations that would not apply to Chicken's pure equilibria, but possibly to its mixed equilibrium (i.e. when both swerve), since this is recognized as a Status quo. This is also how stability is used in the discussion (see for instance Powell 1990, 35; and Zagare and Kilgour 2000, 50-52). In the tabulation (Figure 1.), Chicken is described as yielding an instable depiction; therefore, the colloquial term is used. There are two reasons for this; the first is that there no good synonym for "stability" and while not in keeping with the game theoretic meaning it will harmonize with the more general idea of stability.

Reading the tabulation horizontally, the first aspect in each field is: *speeding hooligans, rational decision making agents* and *two political blocs*. In Russell's model, the actors are not rational, and this is part of Russell's point. That statesmen deliberately choose to play such a game is disconcerting to Russell. In Chicken, the players are rational, they know what they do and they do not make mistakes. Both Russell's model as well as Chicken relates neutrally to the target in the first aspect. Indeed, this is a central part of the crux when representing the Cold War. Were the actors behaving rationally, like in Chicken, or were they not? A significant part of the discussion about Cold War deterrence concerns whether and to which extent the actors in Moscow and Washington were rational. If they were rational, the Cold War could well be thought of as Chicken. To Russell, the mere fact that someone chooses to engage in a game like this signals a lack of rationality. Kahn directly comments on Russell's view and disagrees. To Kahn, Russell's model is a caricature of power politics and that Russell is "too hard on the diplomats", but he concedes that the model contains enough realism to make it plausible. However, insofar as this game is being played it is not being played for some lighthearted reason as "life, liberty and security" may very well depend on playing this game, according to Kahn (Kahn 160, 291). Thus, for Kahn playing the game of Chicken is not irrational; the actors' judgement is not as poor as Russell suggests and the game, when or if it occurs, is important to play.

Schelling also cites Russell and point out that the game of Chicken exists in many versions (Schelling 1966, 116). To Schelling Russell's model is just one of many versions of Chicken – the oldest is described in *The Iliad* – but throughout society the game is played with varying context and versions, which may or may not influence the analysis of the game (Schelling 1966, 117). Hence, to Schelling, Russell's version is not wrong, but also not of primary interest.

In the second part of Hesse's tabulation, which compares the possible outcomes, there is a similarity between the Russell's model and the target. The similarity pertains to one of the actors can swerve, i.e. one of the hooligans can swerve at any point, and one of the two blocs could back down at any given moment. The part of "any given moment" is not modelled in the model Chicken, it only treats the outcomes, but the outcomes as such are represented and are similar to the target. The order of preferences, the third part of the Hesse tabulation, is similar between the two models and the target, i.e. none wants that the cars smash/crash/nuclear war.[10] In research, post the Cold War era the assumption that conflict is the least preferred outcome has been challenged.[11] The issue of whether conflict, in this case nuclear war, is the worst outcome is a central point of discussion in later deterrence modelling. However, as both Russell's model and the game theoretic model Chicken assume that nuclear war is the least preferred outcome, this particular line of analysis can wait.

So what are the consequences of playing such a game? According to Russell, it is a dangerous instable game, largely because the players are not rational. The actors will attempt to win but are not in control of the situation (accidents can occur and someone may be stuck, like Kevin Bacon's character in the film *Footloose*). Since there are no guarantees for such events, it produces a very dangerous game. This is how Russell suggests that the Cold War should be understood. In the model Chicken, the agents are fully rational, and while the game is instable, i.e. one player will always get an advantage and there will be no Status quo, there is always one player who will back down. However, if at least one player is willing to randomize between strategies, the game can result in conflict. While the pure equilibria are suggestive of a game that is quite safe to play, the mixed equilibrium is, like Russell's game, dangerous, as all out conflict can occur. This is however an important difference between the game theoretical model Chicken and Russell's model. While Russell's game is dangerous in a similar way as the mixed equilibrium analysis suggests, Russell also assumes that the players are not rational and are prone to making

---

[10] The idea to defend by threatening with the action, which is one's least preferred, have given rise to several paradoxes both strategic and moral, see for instance Gregory Kavka's account (Kavka 1978).

[11] See in particular Zagare and Kilgour's *Perfect Deterrence* (2000), but also Slantchev (2010) who rely on this assumption and Quackenbush (2011b) who develops this idea.

mistakes.[12] Again, the relationship between the two models and the target is neutral. This is due to the rationality analysis of the actors. In Russell's model the players are not rational, and in Chicken they are. While many thought of the situation of the Cold War as dangerous, if one could trust that Washington and Moscow behaved rationally, the dangerous could be downplayed, more reminiscent of Chicken. Like Kahn, it depends upon how much confidence one has in the diplomats' ability. So what about the stability? This was, and partly still remains, the crucial question – with various answers.

A substantial body of work was dedicated to the analysis of how the stability in a deterrence relationship appeared and connected it to the balance of power between the two blocs. Sometimes this was taken quite literally, like when enumerating the number of nuclear warheads each of the two blocs was in possession of in order to indicate the level of stability. When sufficiently many on both sides this created the sought after stable deterrence balance, sometimes nicknamed MAD (Mutual Assured Destruction) (see for instance Intriligator and Brito 1984). Conversely, when both sides had too few, or when one side had a significantly more than the other did, this implied an imbalance. According to this model, the explanation, why deterrence fails, is that the two players are not equally strong or that they have equal, but insufficient capability. The explanation why deterrence succeeds is that both players have sufficiently high and equal capability. In contrast, the explanation why deterrence fails according to the model Chicken, is that one player always has incentives to yield, when the other attempts an advantage. The explanation why deterrence succeeds is that one of the players will back down in the face of the danger. However, mutual deterrence is not explained – because mutual deterrence does not succeed, at least when analyzed with pure a Nash equilibrium. If explored with mixed Nash equilibrium, the explanation involved has to do with the players being willing to stay in the Status quo when randomizing between strategies with a high degree of probability for one strategy. Unlike the pure equilibria, the mixed equilibrium makes it possible for the other possible outcomes – including conflict. However, this interpretation too is disconnected from the idea of varying capability.

A central issue with the Chicken model is the notion of credibility. In Chicken, the least preferred outcome is the conflict, when neither swerves, the analogy for nuclear war. Hence, threatening with war is not credible since it is known to both players that this is the least preferred outcome, so no rational player would choose this action. Thomas Schelling identified the problem with credibility and expanded the discussion to how a player could credibly commit to defending. To Schelling this is essentially a question of brinkmanship, a competition in risk-

---

[12] Reinhard Selten (1978) developed a version of the Nash equilibrium, which take the fact that a player may make a mistake into account, so-called Trembling hands equilibrium. Also with this type of equilibrium, the Status quo survives.

taking, where the player by showing just how far it was willing to venture near the brink could deter the opponent. In this situation, a move that creates a recognizable risk that may get out of hand will deter an opponent (Schelling 1960, 199).[13] Similarly, Herman Khan contributed ideas of pre-commitment as a method for a player to credibly show that although conflict was the least preferred outcome, the player would still fight. Kahn solves the conundrum of choosing the least preferred outcome by introducing an automated system that triggers a defense in case of attack, thus overriding the preference order, in order to deter credibly. Central to Kahn is also the idea of psychological commitment, i.e. actually preparing society for nuclear war, with bomb shelters and drills, thus signaling commitment to execute a nuclear war (Kahn 1960). Both Kahn's and Schelling's solutions are solutions which recognize the dilemma Chicken depicts, but neither present solutions to the instability of the game itself. Schelling and Kahn treat the equilibria as something to avoid (together with the outcome of conflict) and therefore address the problem of credibility by inventing versions of pre-commitments. They rather seem to view the game as a cautionary tale of a situation that one ought to avoid, something that can be done by various arrangements of pre-commitments (see Quackenbush 2011, 7).

Neither Kahn nor Schelling treat Chicken as a representation of the game that is actually being played between the East and the West all the time, but as a possible instance of it, which should be avoided. Both Schelling and Kahn suggest various bargaining games where prior commitments play a large role, as examples of the game that actually is being played (Schelling 1960, 24; Kahn 1960, 293). Thus, Chicken is a how-possibly explanation, which identifies a certain dynamics from which one can deduce conclusions which will be indicative of a certain policy – but it is not a how actually-explanation.

The neutral relationship between Chicken and the deterrence of the Cold War concerning the stability/instability analysis seems to have driven an important part of the discussion. Authors operating with the hindsight of the early 1990's when the Cold War had ended criticized this discrepancy. To scholars such as Zagare and Kilgour this is a central motivation for why they develop their own deterrence games, which can depict the perceived stability of the Cold War (Zagare and Kilgour 2000). However, Zagare and Kilgour never discuss the mixed equilibrium of Chicken, only the pure equilibria. Similarly, Powell discusses the problem of using Chicken as an analogy for the Cold War, and while Powell recognizes the

---

[13] Schelling suggested a version of Chicken where two players shackled to one and other move close to a brink. If one of the players would start to dance uncontrollably close to the brink this may invoke a risk that would make the other player give in to a demand from the dancer.

mixed equilibrium, the question for him is how it should be understood (Powell 1990, 7).[14] Nevertheless, Chicken remains a much-used model and some scholars champion it as the superior game to understand deterrence (see for instance Walts 1999; and Lawson 2014). It is the neutral relationship between the Chicken model and the target of the Cold War, which seems to make scholars such as Powell, Zagare and Kilgour skeptical about inferring stability, since this is not suggested by the game. Further, they are also dissatisfied with Schelling's and Kahn's suggestions, since neither is a solution to the game. This is also the motivation for all three of them to create other games. It is how credible threats actually work that is their focus.

What seems quite clear from this exposé is that it matters how one interprets the models (Russell's and Chicken) and the target. Zagare and Kilgour seem to think of the Cold War (the target) as dangerous but stable and hence both Russell's analogy and the model Chicken misrepresent. Others seem to have taken a different view. For instance, Robert McNamara, the US Secretary of Defense during the Cuban Missile Crisis, characterized the Cold War as a "Hot War" and that the world "lucked-out" of a nuclear war.[15] This suggests that McNamara rather thought of the Cold War along the lines of Russell's analogy and did not trust the rationality of the players or that they could control all possible contingencies. However, using Chicken as exploiting an aspect of deterrence is still quite possible, and given the many ways in which the situation seems to have been interpreted it lends a perspective that is logically coherent and clear. If the model is viewed as a cautionary tale that one player will always be taken advantage of, and if, like Schelling and Khan, one wants to guard against this possibility, the model reveals important aspects. However, if viewed as a good representation of the Cold War it fails in the central aspect that it does not suggest stability, provided that one views the Cold War as stable. So, when is a model similar enough to make a representation good enough? This clearly has to do with what one's purpose is. If the purpose is to discuss true causes of deterrence as suggestive of stability, Chicken is not necessarily the model one wants. However, if one wants to exploit possible causes it is a model with potential. A more detailed discussion of the purposes of different models can be found in article I.

The issue of credibility has kept deterrence theorists busy. The problem appears clearly if the game is reconfigured as a sequential game where the players take on specific roles. Reinhard Selten, who suggested the game, imagined a firm holding a monopoly that it wants to defend, and a number of entrants, challengers, who want to break the monopoly and enter the market. Hence, one player is a Challenger pondering an attack and one player is a Defender

---

[14] One view is that randomization represents the lack of knowledge (Rubinstein 1991). A second interpretation is that the probability, which the players are willing to play their respective strategies, should be understood as a degree of belief (see Aumann and Brandenburger 1995).
[15] *Fog of War: Eleven lesson from the life of Robert S. McNamara* (2003) Motion Picture. USA: Sony

weighing whether to meet the challenge and defend or to yield. If conflict is still the least preferred outcome, the strategies available to the players change (see article I, for a fuller description of the game). The game corresponds quite well to the basic idea of deterrence "*A refrains from x,* because *B* threatens to do *y".* In the game, commonly known as Entry Deterrence, there are two equilibria, one where the Challenger remains in the Status quo and one where when Challenger attacks and Defender yields. Hence, deterrence will not appear. This is because if Challenger attacks, the rational choice for the Defender is to yield, since yielding is preferred to conflict by Defender. By backwards induction, deterrence cannot occur. However, if, as Selten imagines, that there is a large but finite number of Challengers, then it makes sense to choose conflict for Defender, since this will send a signal to the other Challengers in line, that entering will be met with hostility, thus deterring the other Challengers from entering against the Defender (Selten 1978). This suggests a theoretical weakness in the standard game theoretical solution of the game, which relies on backward induction. In response, Selten developed so-called sub-game perfect equilibrium, which is defined as an equilibrium where every sub-game must be a Nash equilibrium. In subgame perfect equilibrium non-credible threats are eliminated. To prove whether a sequential form game is sub-game perfect is done by backwards induction (Kreps and Wilson 1982).

What do we learn from studying the paradox of Entry Deterrence? The explanandum remains deterrence success/failure. The explanation provided by the game is in this sense not straightforward. The game suggests that deterrence can be a good strategy, but only up to a point; once this point is reached it is better to yield. Hence, to what extent deterrence succeeds hinges on how many rounds the game is played. Entry Deterrence brings out the cut-off point where deterrence ceases to be the dominant solution and where the game theoretic solution, i.e. yielding, suggested by backwards induction, dominates. It is a how-possibly explanation that explores how a defender could reason, but that also points to a problem (deterrence vs. backwards induction). However, unlike Chicken, the situation, or target, depicted by Entry Deterrence is a situation which actually happens, and probably quite frequently. A company dominating a market is likely to think about deterring competition along these lines. For deterrence theorists, the analogy between an incumbent wanting to control a market from challengers who wants to enter it, and a nation-state that wants to deter another nation-state seems to have sparked less discussion than the analogy between hooligans and nuclear war. Partly, this is due to the fact that the analogy between an incumbent with a monopoly deterring many small firms is only partly linked to the state-to-state situation. It is perhaps more rare that one state has to deter many different actors, and therefore the full force of the analogy cannot be

put to use. Moreover, Entry deterrence suggests focus on the technical solution of backwards induction and deterrence credibility, which is of interest, but the problem of credible threats was already known. Although the refinements brought out by studying Entry Deterrence were of interest, it is questionable if the game has not meant more for how games are studied in game theory (see for instance Quackebush's discussion (2011) on the shortcomings of Entry deterrence when studying national defense).

As mentioned in the previous section, Rubinstein discussed different dilemmas that a modeler must face. One of the dilemmas suggested by Rubinstein is the dilemma of absurd conclusions (Rubinstein 2006). This dilemma treats the situation when a model produces or points to an absurdity. In such cases, the question is whether a model construct should be abandoned, or if the absurd conclusion actually reveals something of interest. Entry Deterrence does point to a paradox associated with deterrence. On the one hand, by observing behavior one can see that firms do deter. However, with a standard technique of game theory, backwards induction, this is an irrational way to behave. Further, if one thinks of the game as repetitive, one would actually gain from deterring for several rounds before this strategy becomes suboptimal. This is an absurd result because backwards induction is the method for finding the optimal solution, but it fails in Entry deterrence. Rubinstein, answer to how one should relate to a dilemma of absurd conclusions is that a model is nothing more than a fable or a story. When some of the initial assumptions are applied in a different context, should one be surprised that they do not hold up? Do they make the fable less interesting? In the case of Entry Deterrence the answer is clearly no. For game theorists and economists, Entry Deterrence has been the object of much discussion and further development (see for instance Kreps and Wilson 1982; also Milgrom and Roberts 1982). Again, the purpose of the model matters. If Entry Deterrence is used to explain deterrence, it may leave one wanting, like Chicken. However, if it is to explore a re-occurring conundrum that we associate with deterrence it reveals crucial aspects about deterrence. This suggests that while we learn from the model about the phenomenon of deterrence, it is questionable whether the Entry Deterrence model's primacy does not lie in what it teaches us about certain limits of game theoretic analysis. Entry Deterrence, like Chicken, offers how-possibly explanations of the limits of deterrence and points out the weakness: credible threats. From Entry deterrence, we learn that without credible threats, deterrence is not possible, if at all meaningful.

4. Abandoning the Narrative driven Analogy and Honing in on True Causes

The value of Entry Deterrence to the deterrence theorist seems to have been the sequential move in combination with clearly representing the lack of credible threats. In response, Robert Powell incorporated Schelling's idea of brinkmanship and developed a sequential game. In Powell's game, a Challenger issues a threat and a Defender must decide whether to yield, defend or escalate. If the Defender chooses the latter, Challenger must decide whether it wants to abort, go to conflict or escalate (Powell 1990). The Bayesian game that Powell suggests uses several equilibria, where deterrence is obtained when the defender escalates in a certain way showing its commitment. Hence, the game reveals how threats, which invoke higher risk, can lead to deterrence. However, it does not ultimately change the fact that conflict for both players remains the worst outcome. The ultimate problem is how to make sense of threating with conflict, although that is not in one's interest.

The modelling of credible threats, i.e. modelling how a player could deter rationally, found a solution in Zagare and Kilgour's deterrence games *Perfect Deterrence*. *Perfect Deterrence* presents several deterrence games: a mutual deterrence game, akin to Chicken, but with the possibility of sequential moves (this game is presented in article I), an asymmetric (or unilateral) deterrence game, i.e. a game where one player defends and one challenges (presented in article II), and one model where a third party tries to provide another player with deterrence (so-called Extended deterrence) (Zagare and Kilgour 2000). Zagare's and Kilgour's most central contribution is that they let go of the assumption that conflict is always the least preferred outcome. They use a game with two types of players: one that prefers conflict to yielding and one that prefers yielding to conflict. The players do not know which type they are playing. They end up with a game where symmetric and asymmetric deterrence can be studied under uncertainty, and where updating is possible. The games reveal several equilibria. One central realization is that when both players have a high probability estimation of their opponent's credibility (i.e. prefer conflict to concessions) then there will also be deterrence (equilibrium). It is when the players estimate that their opponent lacks credible (rational) threats that deterrence fails. In this sense, the game also harmonizes with Intriligator and Brito's idea of missile levels (Intriligator and Brito 1984).

Powell, Zagare and Kilgour present games that they believe more accurately capture the core dynamics of deterrence and not special circumstances that lead to dilemmas (Chicken) or to potential paradoxes (Entry Deterrence). Hence, in this respect their critique come from a position where the lack of similarity makes them distrust the representational qualities of the other models. Further, and as discussed in article I, they clearly have an ideal of describing

how deterrence actually functions between nation-states. However, while Powell tends to capture a specific target, the brinkmanship in nuclear deterrence and its connection to credibility, Zagare and Kilgour still seem to have a more generic target in mind. The authors recognize that the context can shift and thus develop games for different scenarios. The search for games that truly represent real targets is apparent with Powel as well as with Zagare and Kilgour.

With ideals towards actual representation, dynamics like dilemmas are no longer central aspects of the game, but rather singular situations that can appear. Both Chicken and Entry Deterrence are important and interesting but point to absurd or strange results. For the deterrence theorists this seems rather to be a problem to overcome than something to contemplate. Few theorists have concluded by inference from Chicken or Entry Deterrence that deterrence as a strategy is paradoxical and that such a defense strategy therefore is absurd. Rather, there seems to be a strong intuition that deterrence is a reasonable strategy, and that the models need to be adjusted to correctly represent reality. The problem of credibility is in this sense solved: either deterrence succeeds because the threat is credible, and it is credible when it is rational. Alternatively, it is credible since on player is willing to venture closer to the brink than the other.

With this shift, the narrative driven analogical reasoning from one source via model to a target is abandoned. Both the Powell model and the Zagare–Kilgour models attempt a more direct approach where inferences are drawn from reality to the model. There are no chickens or entering challengers lurking in the background, only states who want to deter and possibly challenge. Nevertheless, while the narrative may be suppressed, the very modelling is an argument from analogy, where the model functions as an analogy for a particular part of the world. Further, if one believes, as Sugden does, that the strength of a model largely lies in its plausibility, i.e. that its potential to transcend from one specific situation to many potentially interesting situations, then both the Powell and Zagare–Kilgour attempts leave us wanting. The advantage of the narrative driven analogy is that it relates a simple situation that quickly can be understood and applied to many similar situations. With Powell, Zagare and Kilgour we may learn more about the phenomenon of deterrence as a defense strategy, but there is never any suggestion that their models can be extended to discuss other situations.

A different approach is Lisa Carlson's and Raymond Dacey's version of a deterrence model. The authors change the representational theorem of the Perfect Asymmetric Deterrence model from the standard von Neumann-Morgenstern theorem to a prospect theoretical theorem (Carlson and Dacey 2006). The von Neumann–Morgenstern representational theorem shows that a rational player chooses to maximize its expected value of the outcomes of a

given game.[16] The theorem is the underlying choice criterion in game theory. However, results from experimental studies on how people actually make decisions contradict the von Neumann–Morgenstern representational theorem in certain respects. For instance, it has been shown that subjects tend to be more risk-averse when there are large stakes that they could lose and more risk-seeking when there is much to be gained. Daniel Kahneman and Amos Tversky, who pioneered the experimental approach, therefore reweighed the von Neumann–Morgenstern theorem to reflect these results (Kahneman and Tversky 1979). Carlson's and Dacey's deterrence game study how a Challenger's behavior changes depending how it views the value of Status quo, given a modified prospect theoretical function. What makes this deterrence game interesting is that while many deterrence theorists seem to be in the pursuit of a deterrence model which represents true actual causes, Dacey and Carlson rather seem to seek out the how-possibly explanation. By changing the representational theorem for one of the players, the modelers can study how deterrence appears and fails pending on how one player views its current position (Carlson and Dacey 2006). Since the model produces several equilibria, the model suggests possible ways of viewing how deterrence can succeed and fail.

Further, the experimental results are the basis for reformulating the representational theorem of von Neumann and Morgenstern. This means that unlike the original model, the prospect-theoretical theorem actually represents (to the extent that one accepts the results of prospect theory) how real people tend to view risks and prospects. These results contrast all other deterrence models that adhere to the von Neumann–Morgenstern representational theorem. Hence, on the one hand, Carlson and Dacey are searching how-possibly explanations for how a player would behave with a different point of reference; on the other hand, they are reviewing how players with different real dispositions would appraise a given situation.

As noted, the explanations of behavior given by deterrence models tend to be how-possibly explanations. The models explore scenarios, by making some basic assumptions and then following a game theoretic construction and deducing results. Here we have only discussed some of the more central deterrence games, but there are many others (see, for instance, Langlois and Langlois's Deterrence game with bargaining (2006); or Lambertini's nuclear appropriation deterrence game (2013)). Chicken and Entry Deterrence both explore central dynamics, which are of importance to understand. The explanations they yield are how-possibly explanations. They reveal potential difficulties and dilemmas that are important to address if one wants to

---

[16] The von Neumann-Morgenstern representational theorem, also known as the von Neumann-Morgenstern utility theorem, is deduced from four axioms: completeness, transitivity, continuity and independence. The theorem shows that a rational agent will maximize its expected utility if the four axioms are satisfied.

understand deterrence. Other central deterrence models such as Powell's credible deterrence model and Zagare and Kilgour's Perfect Deterrence models tend to attempt how-actually explanations. Although these models may fall short of explaining how deterrence actually works, it is clear that there is a drive for better accounting for the phenomenon by changing the representations to better map the situations of interest. This is principally done in two ways; by stipulating the target more precisely or by changing the model construction. If one were to follow Resnik's view of how-possible explanations as a precursor to actual knowledge, this strive is not only natural but also necessary (Resnik 1991). However, as suggested by those who do not see actual representation as necessary for yielding interesting results, this might not be the avenue that produce the most important result to understand deterrence. Sugden, Rubinstein and Grüne-Yanoff all emphasize that learning and knowledge is obtainable from how-possibly explanations just as from how-actually explanations. The category of how-possibly knowledge is not subordinate to how-actually explanations. From this perspective, there is a strong case for a non-hierarchical view of how-possibly and how-actually explanations. Arguably, models that are more refined and empirically verified may capture actual behavior between two states in a crisis better than Chicken and Entry Deterrence. However, Chicken and Entry Deterrence reveal aspects of deterrence that are central for understanding the fundamental dynamics. From these models, a problem such as credibility was identified and important strategic policies were developed. Furthermore, without the how-possibly knowledge, it is questionable whether the other models could have been constructed, since they are derivatives, developments and refinements of the two original deterrence models. Three of the aforementioned models, Chicken, Entry Deterrence and Perfect Mutual Deterrence are further analyzed in the first paper where they are appraised from different perspectives.

5. Models all the way down

If credibility is central to deterrence and if one wants to investigate whether this indeed is an actual cause for successful deterrence, returning to empirical cases seems to be a way forward. In 2011, Stephen Quackenbush published an article (subsequently expanded into a book), where he used statistical analysis to investigate the predictive power of a deterrence model (Quackenbush 2011a; 2011b). What Quackenbush effectively did was to develop the idea of active dyads (i.e. two rivaling states) to define his cases of deterrence. He then used the outcomes in the Asymmetric Perfect Deterrence game, as proposed by Zagare and Kilgour (1991 and 2000), to investigate to what extent the game could predict historical cases of deterrence situations. Quackenbsuh let the outcomes of the game be the dependent variable and then controlled for

balance of forces between the rivaling states, foreign policy positions between states (so-called S-curves), geographical proximity, democracy and peace years.[17] The results made Quackenbush conclude that the Perfect Deterrence model accurately predicted when deterrence fails and when it succeeds. He then connected this to the idea of credible threats as explored by the model (see Quackenbush 2011a and 2011b).

Quackenbush's research is innovative in that it couples a deterrence model to a statistical model, thereby also circumventing some of the problems that earlier statistical research of deterrence had faced.[18] Based on Quackenbush's findings, should we review the Perfect Asymmetric Deterrence model as a how-actually explanation for how deterrence functions? The theoretical assumptions of the model have been borne out when empirically tested. Thus, it seems like the model identifies true causes for when deterrence fails and when it succeeds. It succeeds when actors have high credibility and it fails when they have low credibility or simply no capacity. The study controls for other factors that otherwise could have explained the results. This means that there is a case to be made that regardless of the how-possibly explanations the Perfect Asymmetric Deterrence game provides, it also explains how deterrence actually works. However, should we be surprised that credible threats can be verified as causal for effective deterrence, if the model makes these very assumptions? Quackenbush needs to make several assumptions and form his data in order for it to fit the model. This does not disqualify Quackenbush's inquiry, but it underscores that also when a model relies on empirical data, the data must also be interpreted and arranged in order to fit the model. Data, like any model, isolate certain features and leave other aspects out. Again, as with a model, this cannot be done arbitrarily; there must be a stringency and methodology in order for the results to be intelligible. Similarly, in article II of this thesis, the Perfect Asymmetric Deterrence model is applied in order to appraise a deterrent situation. The model is instrumental for making sense of a complicated operational situation and it brings clarity in the form of a measure of how and to what extent deterrence took place. However, by relying on the model and looking for relevant data, other aspects are left out. Just as in Quackenbush's article there is a process of isolating and focusing on certain aspects, leaving other structures out of the context.

---

[17] Quckenbush is not the first to have investigated deterrence statistically, some of the more central contributions in this field include Huth and Russett 1984; Lebow and Stein 1990a; Huth and Russett 1988; Lebow and Stein 1990b; and Liberman 1994. Focus of the discussion is on how to separate false positives, situations that appear to be cases of deterrence, but are not (so-called type 1) errors and false negatives, i.e. cases of deterrence that are rejected when they should have been accepted (so-called type 2 errors).

[18] Quackenbush is however, not alone in this type of research. There is similar research by Kim Woosang and Bruce Buena de Mosquita who study crisis bargaining by employing a game model, coding the outcomes and then using multinomial regression analysis. However, one aim of that study was not to investigate the model's ability to predict, but rather to see when and how nation-states make choices in a crisis situation (Woosang and Bueno de Mosquita 1995).

Moreover, if credibility is a true cause for actual deterrence, how should we understand credibility? From game theory, we learn that a credible threat is a rational threat, and a rational threat is a threat that a player has incentives to carry out that threat. However, how does one communicate this commitment to the opponent? Branislav Slantchev focuses on this very question in order to make sense of what credibility actually is (Slantchev 2010). He suggests that there are several approaches, but the most intuitive is that of costly signaling. When an agent makes moves that carry a high cost for the sender, this communicates resolve, since an uncommitted player would not be willing to take on such costs. So what constitutes such costs? The most straightforward way to send a costly signal is by military maneuvers (Slantchev 2010, 30). However, not any military maneuver will do; it must be substantial enough to send the correct signal. The delineation of how credibility can be achieved through costly signals in the form of military maneuvers is interesting and fills an important gap.

In article II, credibility is the explanation for deterrence success and the explanation for the deterrence failure is the lack of credibility. The article also connects this to the military moves and shows how such moves matter to communicating credibility. If credibility is an actual cause for deterrence, and costly signals can explain how this credibility is signaled with military moves, there is a case to be made that from three abstract modelling features, i.e. deterrence, credibility and costly signals, we can arrive at a concrete real world activity such as military moves. However, what type of military move to make is not entirely clear, only that it must be sufficient to reach the desired effect. So, what is sufficiently large? Again, ideals of how-actually explanations end up as how-possibly explanations. The situation that one model seems to build upon another model, which in turn rests on yet another model is recognized by Giere. To Giere each idealization is a model, from arranging data to advanced models that pertain to explain or predict a certain phenomenon. In fact, "[…] scientific reasoning is to a large extent model-based reasoning. It is models almost all the way up and models almost all the way down" (Giere 1999, 42). This is eminently recognizable in the study of deterrence, from deterrence models, to modelling credibility with the help of models of costly signaling all the way down to the representational theorems.

6. Deterrence models, what do they explain?

I have tried to show that the avenue proposed by Schelling and Kahn, which was developed and continued by a range of scholars, has gained significant traction within the research community. The reliance on game theory has been important. It has suggested a methodology and has produced results that have been crucial in the shaping of national defense strategies, most vitally

so during the Cold War. The contact with economics in general and with game theory in particular has meant that methods, important discussions and results are not hammered out in a void, but in communications with other research fields. However, while economists lately have been willing to scrutinize the models they use philosophically, the appetite for philosophical inquiries amongst deterrence theorists has been weak. As this review has shown, there are several aspects related to the deterrence models that would benefit from being discussed from a philosophical point of view. By connecting to the discussion in the philosophy of economics regarding models, several important aspects are discernible. As the previous discussion shows, deterrence models have evolved. While the earlier models were highly idealized and primarily addressed hypothetical scenarios, subsequent models have become more specific, addressing actual situations.

Both Chicken and Entry Deterrence address difficulties associated with deterrence. Chicken shows that in a situation of mutual deterrence there is a dilemma of remaining in the Status Quo and attempting to gain an advantage. In Entry Deterrence, a paradox is explored where a Defender must decide whether to defend against a Challenger or to yield. Both these deterrence games point to certain specific problems associated with deterrence. Hence, as deterrence models, they explain why deterrence fails, not why it can succeed. However, the Cold War was thought of as a stable, albeit dangerous, period, and the models do not bear out that particular fact. As models that explain how deterrence actually happens, Entry Deterrence and Chicken are not necessarily the most revealing, since their explanantia do not explain the explanandum (deterrence success and stability).

The problem of deterrence failure is that it is not rational for the player(s) to confront an attack since conflict is the worst outcome and therefore every other choice dominates the choice of conflict. The explanation for failing deterrence according to Zagare and Kilgour is the lack of credible, i.e. rational threats. An alternative explanation explored by Powell is a different sort of game construction where players can escalate in a certain way, thus deterring a competitor. Both of these solutions are reasonable for showing how deterrence can happen, but is credibility a reasonable solution for deterrence success? If one accepts the assumption that conflict is not the worst outcome, then it probably is. However, is that a reasonable assumption? Some scholars such as Walt find this assumption wrong, although it leads to a depiction of stable deterrence. This is because if one studies nuclear war, conflict should be modelled as the worst outcome, on the basis that no-one would truly like it to occur. Similarly, Powell's brinkmanship deterrence model uses a game form which relies on a specific type of modelling escalation. Consequently, it too uses a quite a specific construction to show how deterrence can succeed.

Both these two advancements largely rely on a how-possibly explanation for coming to terms with deterrence failure. Dacey and Carlson change the representational theorem that is used in game theory and replace it with a Prospect theoretic representational theorem. That theorem is based on experimental results on how individuals tend to view risks and prospects. With this theorem the authors study how deterrence is affected depending on how a Challenger views the Status Quo. Hence, this model too relies on how-possibly explanations, albeit some of its assumptions are based upon how an agent actually see risks and prospects.

It is clear that deterrence theorists have sought to make their models more precise and to address real situations. So much so, that there has been an extensive discussion about which of the deterrence models is best for understanding deterrence. Inferring from our previous discussion, deterrence theorists' view of scientific explanations seems to be akin to Resnik's view of how-possibly explanations as a first step to how-actually explanations. Their opinion seems to be that it is how-actual explanations that are of value. Is this a reasonable attitude? Refining and developing scientific models is an important scientific enterprise; hence, the development in itself is an important process as long as it yields new insights. When viewed as offering a different sort of knowledge it is, however, not clear that the models do provide this sort of development. While Powell, Zagare, Kilgour, Dacey and Carlson certainly make valuable and important contributions, their models also primary provide how-possibly explanations. Contrary to most deterrence theorists' intuitions, this need not be a problem. With how-possibly explanations, much can be learnt. Possible causes can be considered, impossible causes excluded and hypotheses deduced. From the most basic deterrence model Chicken, theorists were able to identify dangers with deterrence as a defense strategy and strategies could be devised which were aimed at avoiding conflict or one actor taking advantage. MAD, flexible response and a hot-line to avoid misunderstanding were all products of a strategic reasoning based on how-possibly explanations. How-possibly explanations provide a heuristic for understanding something like nuclear war for which there is no empirical data to go on. This is indeed the strength of all the aforementioned deterrence models.

As discussed, the empirical validation of the Perfect Asymmetric Deterrence model suggests that this model does capture true causes for when deterrence fail and when deterrence succeeds. This may increase our confidence in the Perfect Asymmetric Deterrence model, and we may therefore accept the premise that for deterrence, credible threats are central. An interpretation in accordance with Resnik's view is that the Perfect Asymmetric Deterrence model has moved our knowledge of deterrence forward, since it provides how-actually explanations of deterrence within a stipulated context. However, if viewing how-possibly and how-actually

explanations as distinct and of different scientific value, it is questionable whether also the Perfect Asymmetric Deterrence model primary provides how-possibly explanations. That it can be corroborated would in this context mean that we could be more confident in its predictive power compared to another model, which has not been tested. This implies that it provides a more fine-grained heuristic, but not that it provides any other type of knowledge than its peers.

From the exposé, it is clear that there is an ideal amongst deterrence theorists to get to actual representation, accounting for true causes. This certainly seems to have been an instrumental ideal, which has provided many new and remarkable developments. However, taking a step back and viewing the entirety of the field of deterrence models suggests that it is not a single model that will ultimately make us understand deterrence; rather it is by applying and considering multitude of them that understanding can be achieved. When this is done, a picture appears where it is possible to see the complexity of deterrence. That the deterrence models systematically and methodologically manage to make sense of some of the more central aspects of the phenomenon is no small feat. The discussion about the game of Chicken, Russell's analogy, and the Cold War reveals that is difficult to understand and interpret deterrence. Russell's analogy and the Roman dictum *si vis pacem, para bellum* only give limited insights.

In contrast, the game theoretic discussion about deterrence has paved way for analyzing the phenomenon in a systematic and detailed way. Sequential moves, credibility, commitment, and costly signals are just some of the modeling contributions borne out by studying the phenomenon of deterrence with the help of game theory. However, such distinctions and insights are not found in one "ultimate" deterrence model, but in the study of the many model versions and in the associated discussions on deterrence. What stands out is the importance of interpreting and understanding how a given model is constructed, what it implies and to what end it has been developed. It is in this realization that we also attain knowledge about deterrence. Models, formal like Chicken, or informal like Russell's proposal, are not only good but also essential when approaching an elusive phenomenon like deterrence. With the aid of a model, certain features can be left out and others amplified or downplayed. This allows for a type of detailed study that otherwise would be difficult – if not impossible. This is evident in article II, where the deterrence model used is a central tool to relate, analyze and explain deterrence. However, the type of knowledge acquired must be interpreted. When this is done and one allows for the fact that deterrence may need representation in different ways, depending on one's end as well as on one's ideals the value of the knowledge yielded will appear more clearly.

It is too late in this already lengthy introduction to say something about the loss of life and destruction of property that well-formed strategies of deterrence can be instrumental in

preventing. However, the mere fact that deterrence can be a crucial force for good, by preventing a war, makes the study worthwhile. However, in so doing it is necessary to be transparent with what type of knowledge one wants and what knowledge one actually gets.

The name of this thesis is *In search for Lost Deterrence*, paraphrasing Marcel Proust's tome *In search for Lost Time*. Proust's narrator reflects on the loss of time and the absence of meaning in the world, only to partly come to terms with memory and meaning in the last volume. Reflecting on the early deterrence models, a common trait is that rather than depicting deterrence, they model the absence of deterrence. With the development of the research field, what was obscured by dilemmas and paradoxes has been complemented with models that better could capture the dynamics of deterrence. However, these models too rely on assumptions and only capture certain deterrence situations. Thus, like for Proust, perhaps the most important realization is that some things will always be elusive and that true meaning will often be indefinite.

Summery Article I: "Comparable Deterrence – Target, Criteria and Purpose"

As mentioned in the introduction there is a debate regarding which deterrence model best represents deterrence. Comparing models is not something that the philosophic or scientific literature has previously engaged with. Rather, as previous discussion has showed the focus has been on representational qualities of a model and ideals guiding model construction. Yet, the deterrence arguments for the advantages with one model over another is reoccurring.[19] As discussed in the introduction, there are various aspects of relevance when appraising a model. Article I 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.

Article I, 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. While the introduction of the thesis has discussed what one can learn from the study of deterrence models, Article I provides a more in-depth analysis to each of these models.

---

[19] This debate primarily consists of Stephen Walt's 1999 article "Rigour and Rigour 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".

Summer Article II: "Deterring the Dauntless – Appraising the effects of naval deterrence against the Somali piracy"

Article II, 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 a model of asymmetric deterrence, the so-called *Perfect Asymmetric Deterrence*, as suggested by Frank Zagare and Marc Kilgour, to understand 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 and the naval counter-piracy activities are measured by the number of performed disrupted pirate vessels. The outcomes are then examined in two models: one which only correlate with the number of warships on station in the two areas of operation (Gulf of Aden and Somali Basin), and one which also 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. The complication of using the model lies with the coding of data and interpreting the results into equilibria. For example, the coding of the two outcomes Status Quo and Conflict suggests that incidents/naval disrupts should be numerically equal. However, this tends to skew all results to the two other outcomes, Defender Concedes and Challenger Defeated. In effect, this would mean that there were almost never periods of Status Quo or Conflict. That would be odd and not in accordance with the general statistical trends. Thus making leeway for some incidents/naval disrupts also during Status Quo and some numerical inequality (given that the amounts are sufficiently high) between incidents/naval disrupts in the Conflict periods gave more reasonable results. The technical elements of the article are discussed in the article itself together with the reliability of the data, which is always a factor when researching a country like Somalia. On a principle level, questions such as data accessibility and how it should be interpreted are potential disablers when studying other similar tactical situations.

While the results in the study are robust, there are several assumptions imbedded in the study. Hence, one must caution not to over-interpret what they mean. Nevertheless, the fact

that the Perfect Asymmetric Deterrence model can be applied to a different situation than it was developed for is interesting since this indicates that it can be put to use to make sense of other types of deterrence situations than it was originally developed for.

References

Aumann, Robert. 1974. "Subjectivity and Correlation in Randomized Strategies", *Journal of Mathematical Economics*, Vol. 1. No. 1: 67–96.

Aumann, Robert and Brandenburg, Adam. 1992. "Epistemic Conditions for Nash Equilibrium", *Econometrica*, Vol. 63. No 5:1161–1180.

Brams, Steven and Kilgour, Marc. 1988. *Game Theory and National Security*. Blackwell Publishing.

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.

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

Giere, Ronald. N. 1999. 2004. "How Models Are Used to Represent Reality", *Philosophy of Science* Vol. 71. No. 5:742–752.

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

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

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

_____ (1988) *Deterrence Failure and Crisis Escalation* International Studies Quarterly 32. No. 1:29–45.

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.

Jeffrey, Richard C. 1983. *The logic of Decision*. Second Edition. University of Chicago Press.

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

Kavka, Gregory S. 1978. "Some Paradoxes of Deterrence" *Journal of Philosophy* Vol. 75, No. 6:285–302.

Kreps, David and Wilson, Robert. 1982. "Sequential Equilibria" *Econometrica*, Vol. 50. No. 4:863–894.

Lambertini, Luca. 2013. "Joining the Nuke Club: A Forward Induction Approach" *Defence and Peace Economics* Vol. 24. No. 1:15–21.

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, Ottawa, Canada.

_____ (1990b) "Deterrence: the Elusive Dependent Variable", *World Politics*, Vol. 42. No 3:336–369.

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

Milgrom, Paul and Roberts, John. 1982. "Predation, Reputation and Entry Deterrence" *Journal of Economic Theory*, Vol. 27. No. 22: 280–312.

Morgan, Mary, S. and Knuutila, Tarja. 2008. "Models and Modelling in Economics"

Myerson, Roger B. 1991. *Game Theory – Analysis of Conflict*. Paperback Edition. Harvard University Press.

Mäki, Uskali. 2009. "MISSing the World – Models as Isolation and Credible Surrogate Systems" *Erkenntnis* Vol. 70. No. 1: 29–43.

Pascal. Blaise. 1662. *Pensées*. Le Livre de Poche.

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.

_____ 2011a. "Deterrence Theory: Where Do We Stand?" *Review of International Studies* Vol. 37. No. 2: 741–762.

_____ 2011b. *Understanding General Deterrence – Theory and Application* Palgrave MacMillan

Resnik, David B. 1991. "How Possibly Explanations in Biology" *Acta Biotheoretica*, Vol. 30. No. 2:141–149.

Rubinstein, Ariel. 1991. "On the Interpretation of Game Theory" *Econometrica*, Vol. 59. No. 4: 909–924.

_____. 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

_____. 1960. *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 Herald and Deising, Paul. 1977. *Conflict among Nations – Bargaining, Decision Making, and System Structure in International Crisis* Princeton University Press

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

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

Woosang, Kim and Bueno de Mosquita, Bruce. 1995. "How Perception 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

_____ 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

I

Routledge
Taylor & Francis Group

Check for updates

Comparable deterrence – target, criteria and purpose

Karl Sörenson[a,b]

[a]Division of Philosophy, Royal Institute of Technology (KTH), Stockholm, Sweden; [b]Department of Military Studies, Swedish National Defence University (FHS), Stockholm, Sweden

ABSTRACT
The last decade has seen several advances in the study of deterrence. These advances have sparked some strong disagreements regarding interpretation of the models and what their contributions signify. This paper appraises the discussions from a model theoretic perspective. It is argued that when comparing rivalling models three aspects; (i) target, (ii) criteria and (iii) type of purpose should be taken into account in order to make a proper appraisal. Informed by these aspects it is evident that the three deterrence models analysed address different aspects, in different ways and to different ends. From this perspective, the so-called Perfect Deterrence model must be recognised as a clear advancement in the research field. Model comparison will always be context relative and a plurality of models should be viewed favourably.

Introduction

Deterrence is back. The Russian annexation of the Crimean Peninsula and the Sino-Japanese dispute over the Senkaku islands are just the latest in a series of geo-strategic events that have prompted many countries to re-think deterrence. From its strategic epicentre for strategic deliberation during the cold war, deterrence as a national defence strategy has after a brief period of focus on small wars over-seas remerged on many countries' defence agendas. The academic research on deterrence has, however, been ongoing, accumulating models to better or further explain deterrence phenomena. With these advances, some strong disagreement regarding how the models should be interpreted and what status they should be awarded has emerged.

This article examines the debate on deterrence models and asks the question: how should we understand a model contribution to a research field such as deterrence? The focus of the debate on deterrence models has been Frank Zagare and Marc Kilgour's Perfect Deterrence. The work re-interprets deterrence with a focus on credibility and information, with results that differ from previous research. Although *Perfect Deterrence* (and the model(s) associated with it) has been viewed as an improvement by some, most notably by Stephen Quackenbush and Branislav Slantchev (see Quackenbush 2011, Slantchev 2011), it also has had its detractors.

CONTACT Karl Sörenson ✉ karl.sorenson@fhs.se

The critique has been broad: ranging from arguing that *Perfect Deterrence* is an inferior depiction of deterrence and that previous models such as the "Game of Chicken" and "Entry Deterrence" are more instructive for policy advice (Lawson 2013), to the model is unnecessarily complicated (see Walt 1999). Hence, on the one hand the research community view *Perfect Deterrence* as an improvement, and on the other hand, it is viewed as wrong, complicated or irrelevant. Neither the defendants nor detractors of *Perfect Deterrence* reference any sort of criteria for why they prefer one particular model to another, where the detractors point out what they find problematic with the model, the defendants tries to show its merits and justify the differences. This sort of argumentation isolates certain elements of the discussion about deterrence models, but tends to miss the overarching question of each model's status and their relation to the research field. Consequently, an unclear discussion has ensued of what any given model contributes or omits. This is unfortunate since the discussion does not settle the larger question of what various deterrence models contribute to the general study of deterrence, nor is it helpful to policy-makers who have to make a decision on which model(s) they should employ to view a particular instance of deterrence. The question seems ultimately to be about what a new model contributes to the study on deterrence. The detractors seem to think of *Perfect Deterrence* as wrong or superfluous, while the defendants seem to take a different approach and view models as complimentary and not necessarily as rivals.

This article's aim is to remedy this situation. Informed by the research on scientific modelling the article investigates the research field of deterrence models and shows that the three models discussed are distinct from one another. Furthermore, it is argued that the choice of one model over the other is justified relative to the context and to what type of understanding one wishes to attain. The article will show that Perfect Deterrence is an important expansion of the research field and improves our understanding of deterrence, but that the choice of any model should ultimately be informed by context and need. For the policy-maker who has to make a decision on how to understand a deterrent situation the model-theoretical approach is instructive as it separates what is being modelled, from how it is modelled and to what end it is modelled, thus enabling comparison and aiding a selection.

Model comparison

The critique of the *Perfect Deterrence Model* (PDM)[1] has mainly centred on three aspects. It has been alleged that the PDM represents deterrence poorly, and that other that more classic deterrence models such as the Game of Chicken and Entry deterrence has more to offer policy advisors (Lawson 2013). Moreover, it is claimed that the model is unnecessarily complicated and that in spite of added complication PDM does not offer any new insight to our understanding of the deterrence phenomena (see Walt 1999).[2] Hence, the critique focuses on: (1) The new model, PDM, does not improve our understanding of the deterrence phenomena; (2) PDM is too complicated/unnecessarily elaborate; (3) Other deterrence models do a better job explaining, advancing research and inform policy than PDM.

What a model depicts, how it is modelled and what purpose it serves are central areas in the discussion on models and these questions are given significant attention in the literature. Roland Giere discusses how a model represents a target, and how this representation relates to the purpose of that representation (Giere 2004). Similarly, Uskali Mäki considers models from two perspectives; ontologically, what their targets represent, and pragmatically, what

the intended purpose with the model is (Mäki 2005, 2009). Michael Weisberg also discusses the model-target relation, which he connects to different representational ideals and criteria for how a model should be constructed (Weisberg 2013). Although there is a general agreement on the importance of these aspects on models, we will need to take the analysis one step further in order to also compare models with each other.

The first question we need to ask is: are the rivalling models, models of the same thing? Several difficulties in the relationship between a model and its intended target have been identified: how a model represents a target and what a model represents. These difficulties are compounded when comparing two models, because not only do we need to ensure the fit between model and target, but we will need to understand how the models differ and if their intended targets differ as well. Thus, it seems reasonable that the model–target relationship must be the first aspect to consider when comparing models, which also connects to the first line of critique (1). Although two models share the same target they might aim at differing features in their target description. As mentioned, what a model's focus is and how this is represented is discussed by Weisberg who suggests that there are different criteria, representational ideals, constraining a model (Weisberg 2013). Models with the same target can represent differently depending on which representational ideals that constrains them. Therefore, a second aspect to consider is the criteria of a given model. This aspect will also be enlightening when considering if the PDM is too complicated, critique (2), since this may relate to the representational ideals of that model. Thirdly, what is the purpose of the model? Given target and criteria, what are the respective purposes of two rivalling models? While target and criteria may be similar, if rivalling models have different purposes it is questionable if they really are to be viewed as rivals at all, but rather as complementary. Purpose, will therefore be the third aspect corresponding to our third (3) point of critique.

(i) Target– What is being modelled?

How does a model relate to its intended target? The relationship between model and target has no fixed procedure or method; rather the modelling literature stresses various aspects of the model–target relationship. Perhaps the most intuitive model–target relationship is the target-direct modelling, where a specific model, e.g. a building and a blueprint of that building, represents a single target. In this case, each part of the model represents a specific part of the target. Modelling state agents require a different approach because not every part of a phenomenon such as deterrence needs to be represented in that kind of model. For instance, the *actual amount* of nuclear warheads, which a state disposes, does not have to be represented; while the *capability* of intercontinental nuclear strikes is something that warrant representation. Thus, representation needs to be understood in a different way.

When modelling a phenomenon Giere suggests that one think of it rather in terms of a system of targets where the modeller decides which part(s) he or she wants to represent. Representing certain causal capacities featured in a target or a target-system, which bears out the aspects one wants to understand, is commonly referred to as idealised modelling (see for instance Morgan and Knuutila 2012 for a discussion on what constitute idealised modelling). With idealisation follows that only certain parts of a target are represented, but as Giere points out this does not mean that we ignore the similarity between a model and target, only that we need to understand it in a different way. While an objective measure of *similarity* between model and target is not required for an idealised model to represent, an

accurate account of causal capacities of the parts of a target one wants to study will still be required for good representation (Giere 2004). Consequently, when considering models which are said to model the same phenomena it is important to differentiate between which part of the target-system or phenomena they mean to represent and what aspects of that particular part of the target they aim to capture. *What* their intended target or target-system actually is and how it is represented will therefore be the first step when comparing models.

(ii) Criteria – How is it modelled?

The second aspect is how the model is constructed. Given a target or target-system, how is it modelled? As already mentioned, Weisberg proposes several criteria for understanding idealised models, for our purposes four of them will suffice; Completeness, Simplicity, A-general and P-general (see Weisberg 2013, p. 105). The Completeness and the Simplicity criteria can be understood as the two most central criteria and one can relate to them as opposites; where what one gains in Completeness one may lose in Simplicity and vice versa in a model (see Mathewson and Weisberg 2009 for a more detailed discussion on the trade-offs in model-building). Completeness consists of two parts; inclusion rules and fidelity-rules. The inclusion rule states that every property of an intended target-system must be included in the model. The model must therefore capture causal and structural properties of the phenomena. The fidelity-rule, in turn, demands of the modeller that every property also is accurately represented. Model completeness is a scientific ideal, which aims to fully account for a phenomenon (Weisberg 2013, p. 107).

Contrasting the Completeness criterion is the Simplicity criterion, which instructs the modeller to include as little as possible in the model, but without infringing on the fidelity rule in the sense that each phenomena characterised is accurately represented. Weisberg points to two justifications for Simplicity. The first is for pedagogical reasons: a simple model helps the scientist to understand data, although the model may in fact distort the data. The second is for construction reasons: when creating a model it is sometimes convenient to start out in a simple fashion and then to build the model bottom up to increasingly handle complexities (Weisberg 2013, p. 109).

A/P-generality relates the number of targets that are represented by the model. A-generality is the *actual* targets that a model applies to, whereas P-generality is the *potential* targets the model could capture. P-generality focuses on a fundamental relationship where the model "abstract away from the actual [target]-system" (Weisberg 2013, p. 110). In contrast, A-generality is when a model has many actual real-world targets. Weisberg's distinction between A/P-generality is subtle, but the generality of a model is a recurring discussion in the scientific community. Robert Sugden seems to be deliberating something similar when he considers the model "market for lemons," a representation of a market for selling used cars (Sugden 2000, p. 5). What is interesting to Sugden is that the "market for lemons" model captures a variety of different situations; labour markets, insurability, and markets in third world countries, while at the same time depicting an actual market for used cars – a very real situation. Hence, there is the actual target: markets for used cars, and the potential targets: insurability situations, labour markets and markets where there is distrust. The example show that while a model can represent both actual and potential targets it may be of greater interest from one of these angels. While the "market for lemons" has high P-generality,

other models may be more interesting for their actual target representation, their A-generality. A model such a "guns s butter," which describe the trade-off between military spending and investments in public goods, is of primary interest for its many actual targets (Dunne *et al*. 2006). Hence, appraising to what extent a model's primary representation is A- or P-generality is of interest since it informs us of which representational ideal it has been constructed with.

(iii) Type of purpose – Why is it modelled?

If a model's target (i) is the answer to *what* a model represents, and the criteria (ii) is the answer to *how* this target is represented it is also important to know *why* the model has be constructed. Kevin Clarke and David Primo suggest that the scientific community at times evaluate models against standards, which are irrelevant given what the model actually was intended for (Clarke and Primo 2007, p. 741).[3]

Clarke and Primo discuss several types of models, which each have different types of purposes. Two of these model types are of interest for the deterrence-model discussion. The first category suggested by the authors is the so-called *Foundational* model. A foundational model does not always reflect real-world situation other than in an abstract way, they typically reference a problem or dilemma. A model such as "Stag hunt," which studies coordination, has proved valuable to philosophers, economists and political scientists ever since Jacques Rousseau first sketched it (Rousseau 1754, p. 39). By alluding to a case specific real situation, the game reveals the general problem how to coordinate in order to receive a greater good than if not coordinating. Hence, a foundational model gives insight to a general set of problems and this is the reason for its development (Clarke and Primo 2007, p. 743).

A second type of model is the so-called *Structural* model. These models aim to organise empirical facts and map these in a model structure that captures key features of a part of research field. Clarke and Primo cites Chris Achen's a model for voting and party identification (Clarke and Primo 2007, p. 744). With his model, Achen generalises eleven recognised empirical findings of voting behaviour, thus structurally summing up a part of the research on voting and party affiliation. Hence, examining the intended purpose of a model will assist in establishing its actual aim. This does not mean that a different application is not allowed, only that evaluating a model in a field it was not developed for and then drawing conclusions about its general appropriateness is misleading.

Align to compare

How the three model-properties (i–iii) relate to one another should in this context be understood as analytical and not confused with the creative process of building a model. For analysing and comparing models it is only important to establish target, criteria and type for each model. It is immaterial which of the aspects one starts with since they are complimentary. Second, similarities and differences between the two models at each given point is considered and explored. By doing this, the parallels and variances between the models are established. The strength with this type of comparison is that it allows one to identify where differences and similarities between rivalling models are located. It also facilities where one might want to put ones focus when comparing.

So far, focus has been on establishing acceptable criteria to study similarities and differences between models, hence predicates of the type "better" and "worse" have been

Player B

	Cooperate C	Defect D
Cooperate C	Status Quo (CC) 3, 3	Advantage B (CD) 2, 4
Defect D	Advantage A (DC) 4, 2	Conflict (DD) 1, 1

(Player A labels the rows)

Figure 1. Mutual deterrence game.

unnecessary. The ideal conditions for model comparison are when two models share target, criteria and type of purpose. When these conditions are satisfied, a comparison in terms of "better" and "worse" will be relatively straightforward. Two structural models with the same target and constructed after the same criteria must also account for all of the phenomena that they aims to represent. Failing to account for phenomena that arguably should be a part of a certain structure should warrant critique.

Target, criteria and type of purpose relate to each other as a form of a checklist, where the aspects are different ways relate to the same form of appraisal. A model can be understood along the lines of its target, a specific set of criteria, or a certain type of purpose. Models that share target, criteria and type of purpose would therefore be ideal for comparison; however, if they differ from one another in any one of these aspects, the comparison will be less straightforward. With this in mind, we now turn to the task of comparing the Classical Deterrence models to the Perfect Deterrence model.

Deterrence models – mapping target, criteria and purpose

(i) Target

The first Classical Deterrence Model (CDM1) assumes two players, or nation states, which both want to deter the other from gaining a strategic advantage, but at the same time, try to attain an advantage for themselves, while making sure full-scale (nuclear) conflict is avoided. Many prominent scholars such as Kahn, Schelling, etc., considered various aspects of this strategic dilemma. The game takes the classical form of a chicken-race; each player has two choices, she can either cooperate (C) or defect (D). If both states co-operate the Status Quo relationship will remain intact (CC). If one player chooses to cooperate but the other chooses to defect the "defector" gains an advantage, outcome (DC) or (CD), if on the other hand, both players defect (DD) conflict breaks out. The preferences appear as ordered pairs in the matrix below. The ordinal ranking for player A is DC > CC > CD > DD and for B is CD > CC > DC > DD (since 4 > 3 > 2 > 1). The game is thus symmetric, i.e. if the players' expected pay-offs are switched they are faced with an identical situation. The game has two pure strategy equilibria, DC and CD (Figure 1).

The game captures a situation when each state makes a decision of cooperating or defecting. Since conflict is the worst outcome in CDM1, choosing to defect would mean that if the

Table 1. CDM1 – target, criteria and type of purpose.

| Name	Target	Criteria	Type of purpose
CDM1	Game of chicken	Simplicity, P-generality	Foundational

counterpart makes the same choice there is conflict. However, given that one can gain an advantage by defecting while the opponent does not, and that this is common knowledge, this would imply that the structure is unstable, and that at times, one player would defect. Viewing the pure strategies (i.e. one defects, one yields) this seems to be contrary to the stable situation that the game was thought to represent. The pure strategies might be seen as unrealistic given the intended target, however, the game's mixed-strategy equilibrium exhibits other characteristics; most importantly is that the status quo survives with a quite high probability, and while conflict can break out it has a relative small likelihood (see for instance Brams and Kilgour 1988). CDM1 is the model that some scholars[4] champion as a better model than PDM for understanding deterrence.

(ii) Criteria

The form of the Game of Chicken (CDM1), has been developed more formally like the version proposed by Steven Brams and Marc Kilgour, but it is often represented as it is here in a more simple form (Brams and Kilgour 1988, p. 45). CDM1 fits the *simplicity* criterion well: it is a minimalistic depiction, which corresponds to the pedagogical aspect of the simplicity criterion. Further, CDM1 conveys the target in the simplest possible way, thus inviting further model development. In fact, it is this very simplicity, which is one of CDM1 major benefits according to some scholars (see for instance Lawson 2013). CDM1 depicts potential situations of an impasse, a dilemma, which potentially can reoccur in many situations. The actual target, two hooligans driving towards each other, is arguably limited. However, the dilemma of swerving/not-swerving, backing-down/not backing down is a reoccurring situation, which is of potential interest to make sense of a range of situations. For instance, an impasse in a negotiation[5] be it political, financial or military can potentially be described in the terms of CDM1. This is the reason why CDM1 also fit the criterion for Weisberg's *potential generality* criterion (P-general), while its actual targets may be few in reality, its potential for capturing a certain type of dilemma remains relevant.

(iii) Type of purpose

CDM1 aims to capture the dilemma situation reminiscent of a crucial aspect of the cold war, however, the interest in the model was not due to the manner in which it organised empirical findings, like a structural model. The keen intent was the dilemma it described, which sparked the point of departure for many deterrence researchers, in spite of its limited target. Herman Kahn uses the model as a departure for his analysis for different types of deterrence (Khan 1960). Thomas Schelling employs it in his study of threats that leave something to chance (Schelling 1960) and Brams and Kilgour expand it with a mixed equilibria-analysis. Robert Powell uses the model to point out why it must be abandoned and replaced with a sequential game (Powell 1990, p. 37) while Zagare and Kilgour considers various features of the model when creating the PDM (Zagare and Kilgour 2000, p. 50). The dilemma has proved to be
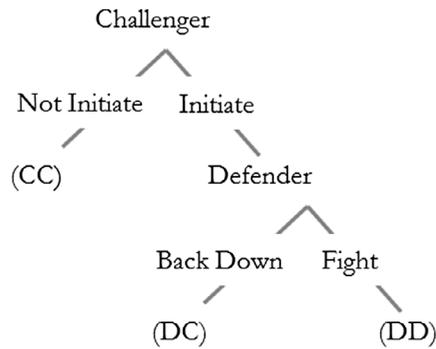
Figure 2. Sequential deterrence.

useful for further analysis. Thus, CDM1 is a foundational model like the cited example of Stag hunt. The research community and policy advisers have embraced CDM1, because of the specific deterrence dilemma it represents even if the real-world phenomena are not very well represented by it. A by-product of this foundational model is the generative effect it has had on other research (Table 1).

(i) Target

In the second *Classic Deterrence model* (CDM2), it is assumed that one of the players contemplates a first move. This makes the game different from CDM1 since one of the players is challenging and one is defending. The Challenger can choose to initiate or to remain in the status quo (CC). If the Challenger initiates then the Defender is left with a choice either to concede (CD) or to go to conflict (DD). The ordinal ranking for Challenger is DC > CC > DD and for B is CC > DC > DD (Figure 2).

   The underlying idea of CDM2 is similar to that of CDM1, where conflict is the worst outcome for both parties. However, representing CDM2 in strategic form, the Defender's best outcome is status quo (CC), while the Challenger's best outcome is (DC). There are two equilibria in CDM2: CC and DC. Although status quo (CC) is an equilibrium, it is not subgame perfect since if Challenger decides to initiate, Defender's best response is to concede (DC) since DC > DD.

   In CDM2, there is a sequence of play, with a first mover (the Challenger) and a Defender. The CDM2 cover many situations and is sometimes referred to as "Entry Deterrence." CDM2 captures a peculiar phenomenon with deterrence: given the payoffs, it is irrational for Defender to defend since this will lead to conflict that is more costly than simply yielding. Still, if known to Challenger it will always be able to get his preferred outcome by challenging and Defender will concede. Deterrence will therefore never succeed. When modelled in this manner the model captures a fundamental and important problem with deterrence. This problem has been formulated as a paradox by Reinhard Selten, and it has since then been the object of discussion and further model development.[6] What is crucial with CDM2 in its original form with complete information is that it reveals the importance of sequence of play and as a consequence the importance of commitment for credible deterrence.

Table 2. CDM2 – target, criteria and type of purpose.

| Name	Target	Criteria	Type of purpose
CDM2	Entry deterrence	Simplicity A-generality	Foundational

(ii) Criterion

CDM2, as represented here, is used as a point of departure by many scholars. CDM2 also captures a dynamic, which will reoccur between an incumbent and a challenger. Every firm that wants to break into a new market will have to consider the problem of entry, and every firm, which controls a market, will have to decide if it should confront or allow a competing firm. Similarly, every state considering challenging another state or a state that is facing a challenge must decide whether to allow or face the challenge. Entry deterrence is an actual problem. This suggests that the game has an *actual generality* (A-generality) and seem to have been developed with this generic situation in mind. As in the case with CDM1, CDM2 uses few assumptions to target a specific situation. The model is pedagogical persuasive and, just as in the case with CDM1, it has been used for further model-development from its rudimentary original form. Hence, like CDM1 it corresponds to the *simplicity criterion*.

(iii) Type of purpose

CDM2 has been studied with regard to the paradox associated with it, by Reinhard Selten, but also by Robert Powell who relates it to credibility, and by Eric Rasmussen, in relation to the role information may play (Selten 1978, Powell 1990, and Rasmusen 2007, p. 160). However, regardless of the level of sophistication of the analysis developed by the authors, they all depart from a game as described above. The keen interest in the model has to do with a foundational problem with deterrence. In this sense CDM2 and CDM1 are similar in that they focus attention on an imbedded problem with the concept of deterrence. While the problem of deterrence can be analysed in different ways, understanding Entry Deterrence and the problem it represents is important in order to understand a fundamental part of the deterrence phenomena. Hence, CDM2 will remain important in deterrence theory and in game theory since its type of purpose is foundational (Table 2).

(i) Target

In the Perfect Deterrence model (PDM), mutual deterrence takes the following form: just as before player A and B have the choices of either cooperating or defecting, and like before there are four possible outcomes: Status Quo, Advantage A, Advantage B and Conflict. The preference between the outcomes are however different. For A: $a_{DC} > a_{CC} > [a_{DD}$ and $a_{CD}]$ and for B: $b_{CD} > b_{CC} > [b_{DD}$ and $b_{DC}]$, hence the Conflict and Advantage for the opponent are left open. Furthermore, each player's estimate of the other player's preference of Conflict over Concession is taken as a measure of the other player's credibility (Zagare and Kilgour 2000). The preferences a player may or may not have determine which types of player it is: A is considered of the Hard type when A: $a_{DC} > a_{CC} > a_{DD} > a_{CD}$ and soft when A: $a_{DC} > a_{CC} > a_{CD} > a_{DD}$, conversely B is considered Hard when A: $b_{CD} > b_{CC} > b_{DD} > a_{DC}$ and soft when B: $b_{CD} > b_{CC} > b_{DC}$ and b_{DD}.
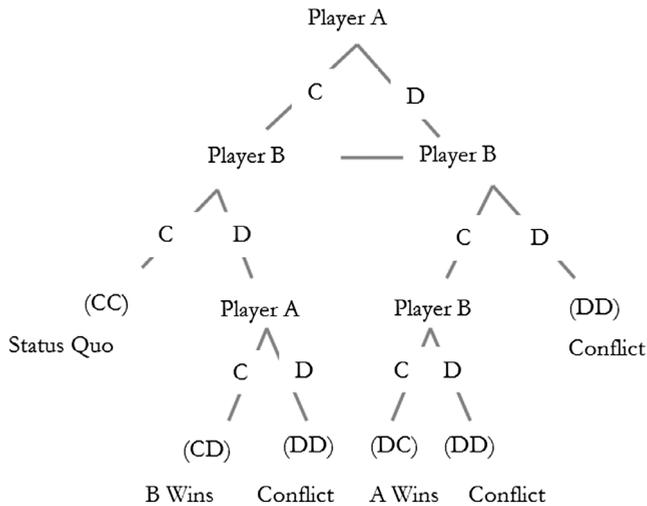
Figure 3. Perfect deterrence mutual deterrence game.

In order to model the uncertainty about the player's preferences the model introduces two variables. Player A is Hard i.e. $[a_{DD} > a_{CD}]$ with the probability p_A and Soft $[a_{CD} > a_{DD}]$ with the probability $1-p_A$. Similarly, Player B is Hard i.e. $[b_{DD} > b_{DC}]$ with the probability p_B and Soft $[b_{DC} > b_{DD}]$ with the probability $1-p_B$. With complete information of the game there are four possible outcomes, see Figure 3. However, the only time when deterrence succeeds is when the players are credible, i.e. Hard, and this is the only time when deterrence is a sub-game perfect equilibrium. The name Perfect Deterrence theory is derived from the sub-game perfect play. In reality, conditions of complete information rarely exists, hence the game under incomplete information is central, and is what Zagare and Kilgour have focused on. Under incomplete information there are eight possible equilibria, which can be categorised in three groups, defined by the thresholds; three deterrence equilibria (called Sure-Thing, Separating and Hybrid Deterrence), three Attack equilibria for player A and three for player B, and one Bluff Equilibrium (Zagare and Kilgour 2000). Given the probability distributions, the game can become a Game of Chicken, but this is just one of many possibilities. It all depends on the preferences of the players and how credible they believe their opponent to be, i.e. which of the many versions of the game that is actually being played.

PDM is in many ways a different account of deterrence than CDM1 and CDM2; it relaxes the assumption about worst outcome for the players, which expands its target-system compared to previous modes. However, it also makes several assumptions, which CDM1 and CDM2 do not: it tries to capture the credibility of the player's threats, and for the possibility for Challenger to update its choice after Defender has made his, making the game variable. This potentially narrows the target-system the model is meant to represent. With the same assumptions, it also aims to model nuclear and conventional mutual deterrence, asymmetric deterrence situation as well as extended deterrence situations. Although the latter two are not considered here, they are built on the same understanding of credibility and Bayesian updating. Hence, from a *target* perspective (i) the PDM opens up for rational play (without a paradox or dilemma), capturing credibility and updating which expands the intended target into a system of targets.

Table 3. PDM – target, criteria and type of purpose.

Name	Target	Criteria	Type of purpose
PDM	Mutual Deterrence w/without credible threat Challenger's second choice	Completeness A-general	Structural

Table 4. Comparison of deterrence models.

	CDM1	CDM2	PDM
(i) Target	Game of chicken dilemma	General entry deterrence	Mutual deterrence, unilateral deterrence, extended deterrence relating credibility of threats
(ii) Criteria	Simplicity, P-general	Simplicity, A-general	Completeness, A-general
(iii) Type	Foundational	Foundational	Structural

(ii) Criteria

In the PDM the Bayesian model intended to isolate different types of players allowing several equilibria enabling more sophisticated, but also more complex results. The PDM examines the implications of variable credibility for the players and the possibility for Challenger to make a second decision after having initiated with updated information on Defender's reaction to a challenge (node 3). This gives PDM many *actual* targets, since a Challenger and a Defender would in most situations have to estimate the credibility of its opponent. Similarly, the possibility to abort an attack is something that a Challenger will have as an option after having initiated a challenge (A-generality). The variable credibility and the second decision for Challenger require more in terms of representation. Given the critique of PDM's complicated structure it seems that it is exactly the representations that also makes PDM fall short of pedagogical persuasiveness – a requirement for the simplicity criterion (see Walt 1999; and Lawson 2013). Hence, the completeness criterion seems to apply to PDM to a greater extent than to CDM1 and CDM2. On the other hand, other scholars have used PDM for further development, which would suggest that while less simple than CDM1 and CDM2, PDM is not insurmountably complicated (see the end-discussion for some examples). PDM is an inclusive model as it models a second move for the Challenger and a models variable credibility, and hence, the Completeness ideal is noticeable.

(iii) Type of purpose

The PDM systematically goes through various versions of deterrent relationships and is, throughout, in communication with the research field. By systematically mapping many deterrence situations and referencing research, Zagare and Kilgour seem to have a deterrence model in mind which is structural and which can account for many different deterrence situations. Interestingly, in a study from 2010, Steven Quackenbush re-interprets the equilibria of the Unilateral PDM to predict outcomes of (would be) deterrent relationships, which is tested with the help of dyads (Quackenbush 2010). The Quackenbush study could be seen as a test of how well the structure of PDM represents empirical situations of deterrence (Table 3).

Appraising deterrence models

Summing up the discussion regarding the three models we can conclude the following (see Table 4): While CDM1 was modelled after one, albeit global, conflict, its target (i) is uniquely outlined. It aims to represent an important strategic aspect of the cold war, a dilemma. The model's simplicity is arguably one of its strengths; a strength which when paired with the portrayed dilemma situation has sparked a lot of creativity. The target of CDM2 is different from CDM1, although it idealises its target, it also assigns fixed roles of Challenger and Defender and study this particular dynamics in sequence as an extensive form game. While the targets of CDM1 and CDM2 differ, they share traits about criteria and type. Both are simple in that they only model certain aspects, but while CDM1 potentially is generic to dilemma situations (P-General), CDM2 is applicable to many actual situations (A-General). This is mirrored in their type of purpose, which is foundational in that they strive to capture fundamental (albeit different) characteristics of deterrence.

The PDM was introduced out of perceived deficiencies with the CDM1 and CDM2, and consequently there are differences. The PDM is more encompassing than CDM1 and CDM2 in that the game of chicken can be expressed, as well as basic entry deterrence, but it also captures several other instances of deterrence, from mutual deterrence, conventional or nuclear, with or without sequence of play, and while not considered here, also unilateral deterrence and extended deterrence. More importantly, it studies the credibility of the players and leaves it open if conflict is preferred over concession or vice versa. Hence, the target of PDM covers CDM1 and CDM2. PDM's larger target necessitates certain steps when modelling it (such as the Bayesian updating) in order to represent the intended target accurately. This more complete ambition is a consequence of its wider target of deterrence. This infringes on how easy PDM is to understand compared to CDM1 and CDM2, which both focus on distinct aspects of deterrence and needs less in order to represent. The wider target, the ambition of completeness is reflected in its type of purpose, which is a structural account of many different deterrence relationships.

Aligning the CDM1, CDM2 and PDM there are some instances where they overlap. For instance, at a certain moment CDM1 and PDM overlap if PDM is viewed as representing a situation where both players are contemplating a first move simultaneously. One possible answer is that since PDM covers CDM1, one could simply choose this model. However, if it is only the dilemma one is interested in, i.e. one assumes that neither contender prefers conflict to concession, then the PDM could be unnecessarily complex. Since that dilemma is the target of the CDM1 and the purpose is to study this particular situation, CDM1 would be the appropriate choice. The relatively more complex PDM might even obscure that particular situation. On the other hand, if there is a need to understand how this relationship may change and how credibility plays a role, then the PDM has more to offer. What this short example show is that simply having a target in mind is insufficient ground for model selection; the type of purpose and criteria needs also to be taken into account.

As mentioned in the beginning of the article, there are different views vis-à-vis the various deterrence models one can take. While the critique directed against PDM takes the position that the model is redundant since it does not add anything new, in addition to being complicated, the defenders seem to view models in general as complimentary. However, in certain instances, or certain well-stipulated contexts, it is possible to think of one model as better than another. Given the importance of differentiating between a model's target,

criteria and type of purpose, the first view seems to be problematic for two reasons. First, target expansion provides new insights. What is of primary interest is not where two models overlap – it is where they differ. It is what is different with PDM to CDM1 or CDM2, which is of interest. What PDM adds, for instance in its analysis of credibility, is of interest to research on deterrence. Second, model rivalry is an important issue, but that is not the main purpose of the PDM-CDM1-CDM2-comparison. As can be noted from our discussion, target, criteria and type of purpose differs quite substantially between the three models. Hence, the detractors of PDM take an uniformed view on what the general scope and limits of scientific modelling are. The defenders of PDM who has a more inclusive view of models, has therefore also a more reasonable view. When introducing a new model to a general field of study, provided it does not misrepresent, it should be thought of as an expansion of a research field rather than a constraint. It is what the deterrence models inform us collectively, which is of interest to the general study of deterrence – and this is what some of the debate on the *Perfect Deterrence* has misconstrued.

The last point is important. Since the publication of *Perfect Deterrence,* many deterrence theorist have made other contributions to the field. For instance, Branislav Slantchev has investigated the relation between costly moves to deterrence stability (Slantchev 2011). Slantchev's focus is on what constitutes credible commitment and uses the notion of credibility as proposed by Zagare and Kilgour. The purpose of the deterrence models proposed by Slantchev tend towards the structural type as he explores causal connections between military and costly moves. In a related, yet different vein, Lisa Carlson and Dacey have investigated the implications of prospect theoretical results for deterrence stability by studying a Challenger's perception of the Status Quo (Carlson and Dacey 2006). That study makes reference to the original Zagare and Kilgour model, but reweighs the payoff function in order to study how a Challenger values the Status Quo over the possibility of going to conflict. Since, Carlson and Darcy rely on empirical findings on risk-aversion and risk-seeking behaviour, its purpose seems to be of structural nature. The aforementioned research projects are but two of many which use the vast span of deterrence modelling to further investigate the same phenomena.

It was suggested in the beginning of this article that deterrence as a defence policy has been reinstated as the dominant defence strategy for many state actors. Looking ahead it seems likely that the study of deterrence will experience a renaissance. Given the target, criteria and type of the deterrence models here reviewed it is probable that the aspects of these models will form part of further model development and application. For instance, given the lack of defence displayed in securing parts of the Ukrainian territory, more narrowly formulated deterrence models, which explicate the factors that lead small states to fail would be of acute interest, and so is an expanded studies of defence alliances. Hence, narrowing the target and changing the purpose to more explicative models, would deepen our understanding of the deterrence phenomena. Similarly, there is a gap in the deterrence literature, already identified by Lawrence Freedman, which concern the parallel research on deterrence in the field of criminology. Criminologists have for long studied deterrence, but the international security literature and criminology literature still have not caught up with one another (Freedman 2004, p. 60). A widening of the deterrence target to include military operations and tactics should be a welcome addition. In this context models explicating and predicting deterrence behaviour would be an interesting development of the deterrence field.

Conclusion

In this article, I have argued that the deterrence theoretical discourse on model comparison can be made more intelligible if informed by insights from the research on scientific models. When considering the deterrence debate from a target, criteria and type of purpose point of view, it becomes clear that the three referenced deterrence models; Game of Chicken, Entry Deterrence and Perfect Deterrence, represent different targets, are represented in different ways and to different ends. The field of deterrence should not be viewed as a horse-race between competing models; it is a field of communicating vessels, where each model is adding, subtracting, tweaking and re-making previous work. However, given a specific context, target, criteria and type of purpose will be instrumental to decide which model will be most beneficial to use.

The three aspects, introduced in this article, allow in an approximate way to choose between different models in order to better select the one(s) that informs the most, given a particular context of interest. In this perspective the benefits with PDM do not supersede or replace previous deterrence models such as CDM1 and CDM2; but instead relays other important aspects of the phenomena known as deterrence, which given a specific context can be crucial for our understanding.

Notes

1. For matters of convenience, I adopt the standard of identifying the Classical Deterrence Theory models as CDM, with appropriate index and the models associated with the Perfect Deterrence Theory as PDM.
2. These contributions are part of a debate, which seems to reappear with regular intervals, for the more central contributions see Stephen Walt's 1999 article "Rigour and Rigour Mortis – Rational Choice and Security Studies" and Frank Zagare's reply in the same year "All Rigour No Mortis" ,Stephen Quackenbush's discussion from 2011 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 1999, 2013 "Deterrence Then and Now – There Is No Going Back".
3. Clarke and Primo discusses five types of purposes a model can have; foundational, structural, generative, explicative and predictive (Clarke and Primo 2007). The motivation for examining different models' intended purpose is that many social scientist only values prediction and tend to value a model on after its ability to predict, when in fact models can have a variety of different types of purposes. While not considered here, the clearest case of one model being superior to another is the predictive model. For example, two models predicting the infection rate of Ebola might have the same target (the rate of how Ebola spreads), and the same criteria (e.g. Simplicity and A-general) and type of purpose, (to predict the spread of Ebola). In this context, it can make sense to discard one model because of its inferiority to predict compared to its rival (Kupferschmidt 2014).
4. See, Lawson, Fred. 2013. "Back to the Future in the Study of Deterrence" *St Anthony's International Review* Vol. 9 No. 1:144–156.
5. A contemporary example of this is the Greek debt negation with the EU, which was described in the media as a Game of Chicken (Guardian 20 Feb. 2015).
6. See for instance Paul Milgrom and John Roberts article "Predation, Reputation and Entry Deterrence" (Milgrom and Roberts 1982) and David Kreps and Roberts Wilson's article "Sequential Equilibria" (Kreps and Wilson 1982) who further discusses Selten's idea of Entry Deterrence.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

Karl Sörenson is a PhD candidate at the Royal Institute of Technology (KTH), Division of Philosophy and the Swedish National Defence University (FHS), Department of Military Studies. His thesis explores deterrence from a game and decision theoretic perspective. Before his current position, Sörenson worked at the Swedish Defence Research Agency (FOI) as a senior analyst.

References

Brams, S. and Kilgour, M., 1988. Game theory and national security. Blackwell Publishing

Carlson, L. and Dacey, R., 2006. Sequential analysis of deterrence games with a declining status quo. *Conflict management and peace science*, 23 (2), 181–198.

Clarke, K. and Primo, D., 2007. Modernizing political science: a model-based approach. *Perspectives on politics*, 5 (4), 741–753.

Dunne, P.J., Smith R.P., Willenbockel D. 2006. Models of military expenditure and growth: a critical review. *Defence and peace economics*, 16 (6), 449–461.

Enderlein, H., 20 Feb 2015. It is up to Germany to end the game of chicken with Greece. *Guardian*. Available from: http://www.theguardian.com/commentisfree/2015/feb/20/greece-germany-eu-bailout [Accessed 23 Feb 2015].

Freedman, L., 2004. *Deterrence*. Malden, MA: Polity Press.

Giere, R., 2004. How models are used to represent reality. *Philosophy of science*. 71 (5), 742–752.

Khan, H., 1960. *On thermonuclear war*. Princeton University Press, reprinted by Transaction Publishers 2007.

Kreps, D. and Wilson, R., 1982. Sequential equilibria. *Econometrica*, 50 (4), 863–894.

Kupferschmidt, K., 2 Dec 2014. Star statistician Hans Rosling takes on Ebola. *Science insider*. Available from: http://news.sciencemag.org/africa/2014/12/star-statistician-hans-rosling-takes-ebola [Accessed 20 Feb 2015].

Lawson, F., 2013. Back to the future in the study of deterrence. *St Anthony's international review*, 9 (1), 144–156.

Mathewson, J. and Weisberg, M., 2009. The structure of tradeoffs in model building. *Synthese*, 170 (1), 169–190.

Mäki, U., 2005. Models are experiments. experiments are models. *Journal of economic methodology*, 12 (2), 303–315.

Mäki, U., 2009. MISSing the world. Models as isolations and credible surrogate systems. *Erkenntnis*, 70 (1), 29–43.

Milgrom, P. and Roberts, J., 1982. Predation, reputation and entry deterrence. *Journal of economic theory*, 27 (22), 280–312.

Morgan, Mary S. and Knuuttila, Tarja, 2012. Models and modelling in economics. In: Mäki, and Uskali, eds. *Philosophy of Economics. Handbook of the philosophy of science*. Elsevier, pp. 49–87.

Powell, R., 1990. *Nuclear deterrence – the search for credibility*. Cambridge: Cambridge University Press.

Quackenbush, S.L. 2010. General deterrence and international conflict: testing perfect deterrence theory. *International interactions*. 36: 60-85.

Quackenbush, S.L. 2011. Deterrence theory: where do we stand? *Review of International Studies*, 37 (2), 741–762.

Rasmusen, Eric, 2007. *Games and Information*. 4th ed. Blackwell Publishing.

Rousseau, Jean-Jacques, 1754. Discours sur l'origine et les fondements de l'inégalité parmi les hommes. Le Livre de Poche, 1996.

Schelling, T.C., 1960. *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Selten, R., 1978. The chain store paradox. *Theory and decision*, 9 (2), 127–159.

Slantchev, B.L., 2011. *Military threats – the costs of coercion and the price of peace*. Cambridge: Cambridge University Press.

Sugden, R., 2000. Credible worlds: the status of theoretical models in economics. *Journal of economic methodology*, 77 (1), 1–31.

Walt, S., 1999. Rigor or rigor mortis? Rational choice and security studies. *International security*, 23 (4), 5–48.

Weisberg, M., 2013. *Simulation and similarity – using models to understand the world*. New York: Oxford University Press.

Zagare, F., 1999. All rigor, no mortis. *International Security*, 24 (2), 107–114.

Zagare, F., 2013. Deterrence theory then and now: there is no going back. *St-Anthony's international review*, 9 (1), 157–167.

Zagare, F. and Kilgour, M., 2000. *Perfect deterrence*. Cambridge: Cambridge University Press.

II

CrossMark

ARTICLE

Deterring the dauntless—appraising the effects of naval deterrence against the Somali piracy

Karl Sörenson[1]

Abstract This article considers whether the Somali pirates were deterred by the naval task forces between the years 2009 and 2013. By disaggregating data and using previously unpublished records regarding the naval operations, two areas of operations are identified as potential periods of deterrence. The article uses a model of asymmetric deterrence to study the outcomes and equilibria of the navy-pirate interaction. It is found that the naval operations eventually 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 operational focus on the Gulf of Aden coupled with the fact that the area is relatively smaller than the Somali Basin enabled the naval credibility, thus effectively denying the pirates access to the sea. Conversely, limited attention by the naval units and the long Somali southern coast with its open waters impeded naval control in the Somali Basin. In connection to these findings some conclusions regarding naval deterrence are discussed.

Keywords Deterrence · Piracy · Somalia · Naval operation

1 The decline of Somali piracy

In the beginning of 2009, the international community launched a series of naval operations to combat the Somali piracy. The expressed task was to "detect, disrupt and deter" the Somali piracy (EU NAVFOR Operation Atalanta Tasking). After four years of on-going naval operations, the piracy decreased significantly to almost no recorded attacks during 2013. Were the Somali pirates deterred by the naval task

✉ Karl Sörenson
  karl.sorenson@fhs.se

[1] Royal Institute of Technology (KTH) and Swedish National Defence University (FHS), Drottning Kristinas väg 37, Box 27805, 115 93 Stockholm, Sweden

🖄 Springer

groups? Moreover, if they were how can we understand this deterrent? These are the two questions examined in this paper.

The waters off the coast of Somalia have been treacherous for many decades. From the time organisations such as the International Maritime Organisation (IMO) and the International Maritime Bureau (ICC-IMB) has kept records of piracy and armed robbery at sea there have been reports of acts of piracy associated with Somalia. During the late 1990s, the piracy intensified and in 2002, the pirates started to hijack ships and demand ransoms. The practise of hijacking ships continued to increase and saw a record numbers during 2005 (see for instance Hansen 2009 on the origins of the Somali piracy). During 2007, France together with some other nations launched a naval operation to ensure protection of the World Food Programme's (WFP) ships bound for Somalia and to hunt pirate vessels. The operation, known as Alcyon, managed to secure the WFP ships, but was insufficient to tackle the broader piracy problem. In 2009, the EU NAVFOR Operation Atalanta took over the responsibilities of Alcyon. Shortly after, two other counter-piracy operations, the NATO Operation Ocean Shield and the Coalition of Maritime Forces (CMF) were also deployed. At its peak, the counter-piracy operations encompassed more than 20 naval units on station (Sörenson and Widen 2014). In spite of the naval operations' efforts, the Somali piracy continued to increase and reached a new record number of 49 hijacked ships during 2010. The naval operations and the international community increased their efforts to combat the Somali piracy, but also to better tackle the broader security problem in Somalia. After an intensive period, the Somali piracy started to ebb out and finally seemed to have ceased with almost no recorded incidents during 2013 (see ICC-IMB for statistics).

There are several factors, which may have influenced the decline of the Somali piracy. The research has mainly focused on three factors: (i) the naval intervention, (ii) the shipping community's own security precautions and (iii) the stabilisation of Somali society. An important article on the subject, by Anja Shortland and Sarah Percy, study the decline in pirate activity linked to Somali economy and naval intervention (i.e. factor (i) and (iii)). The Shortland-Percy paper, which study the period 2002–2010, conclude that while anarchic condition does not benefit piracy, weak society and corruptible governance does (Shortland and Percy 2013). In connection, the authors suggest that the naval intervention in 2009 displaced the pirates from the Gulf of Aden (GoA) into the Somali Basin (SB). A similar study by Ryan Jablonski and Steven Oliver also explores economic opportunity in Somalia, and as a part of the study consider the naval deterrent, i.e. factors (i) and (iii) (Jablonski and Oliver 2013). The study concludes that much remains unknown of the drivers behind the Somali piracy and that the lack of quantifiable data makes it difficult to measure an eventual deterrent effect the naval operation might have had. In another study, Anja Shortland and Federico Verese investigate to which extent the local societies where willing to protect the Somali pirates during its heydays (factor (iii)). The authors conclude that a shift in the local population's attitude given an improvement of the economic situation (factor (iii)) contributed to the decline of the Somali piracy (Shortland and Varese 2014). Karl Sörenson and J.J. Widen consider the tactical interaction between pirates, navies and shipping (i.e. factors (i and ii)). They conclude that there seems to be three distinct periods in the exchange between the Somali pirates, the naval operations and the shipping community (Sörenson and Widen 2014). While these aforementioned studies are quite specific concerning factors and their relation to the Somali piracy, there is a

substantial body of research, which focuses only on factor (iii). Best known is perhaps the works of Ken Menkhaus [1] and Stig Jarle Hansen. [2] Hence, when reviewing the literature, although factor (iii) dominates the research field, the other two factors have also been given due attention, and while never an explicit focus, the naval deterrence figure in the background of many of these studies as a possibly important aspect of the general decline of the Somali piracy.

This article adds to previous studies by exclusively studying the naval deterrence (i), and with the help of factor (ii) and (iii) say something more definite about the navies' impact. In order to understand if the naval task groups deterred the Somali pirates, a deterrence model is employed to study the outcomes of the navy-pirate interaction. The outcomes are then analysed as equilibria defined by the deterrence model. The article concludes that while the naval intervention is the best explanation for the outcomes in both areas of operations it was only in the Gulf of Aden that the naval task forces managed to deter the Somali Pirates. In the Somali Basin, the deterrence objective was however, not met. The reason for the difference between the two areas is that the naval operations prioritised its units to the Gulf of Aden, and as these waters are more narrow than the water off the coast of southern Somalia, the navies eventually managed to deny the Somali pirates access. Consequently, the Somali pirates came to regard the navies patrolling the Gulf of Aden as a credible threat. In the Somali Basin, this effect never materialised, despite an intensified effort by the navies towards 2012 to increase their patrolling in the area.
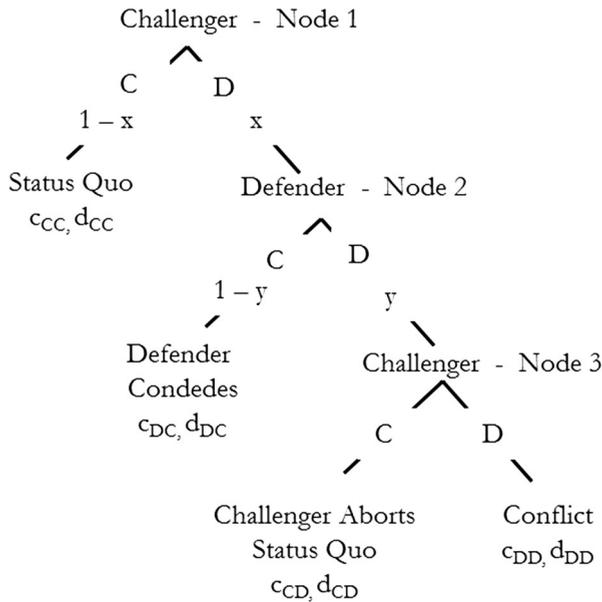
2 Deterrence

In order to study the behaviour of the Somali pirates and the naval operations structurally, a game theoretical model is used. The model, introduced by Frank Zagare and Marc Kilgour model in 1993, entitled 'Asymmetric Deterrence' (Zagare and Kilgour 1993) explores a deterrent situation in form of a dynamic game with variable credibility. The challenger-defender interaction explored by Zagare and Kilgour can be recognised from other deterrence games such as entry deterrence (Rasmusen 2007). The game takes the following form: a *Challenger* decides whether to initiate or not to initiate a challenge. If it does not initiate, the game ends in the status quo (or CC in Fig. 1). If, on the other hand, the Challenger decides to initiate, the *Defender* must decide whether to allow it or to resist the challenge. If the Defender chooses the former alternative, the game ends and the Challenger gains an advantage (DC). If the defender decides to meet the challenge, the Challenger must in turn decide to either back down and give the Defender an advantage (CD) or go to conflict (DD). The extensive form of the game is presented in Fig. 1.

The Challenger strictly prefers an advantage (DC) to the status quo (CC). Secondly, each player prefers winning without a confrontation, (CD) for the defender and (DC) for challenger, or the status quo (CC), to conflict (DD). Finally, each player knows the other player's preferences satisfy these restrictions, but is uncertain about whether the opponent prefers conflict, (DD), to capitulation, (CD) for the Challenger and (DC) for Defender. The

---

[1] See for example Menkhaus, Ken. 2007. 'Governance without Government in Somalia' International Security, vol. 31, no. 3
[2] See for example Hansen, Stig Jarle. 2009. Piracy in the Greater Gulf of Aden Norwegian Institute for Urban Regional Research, NIBR 2009:29

Fig. 1 Asymmetric deterrence model

players' preference orders decide which type of player it is. The Challenger is of the hard type when: $c_{DC} > c_{CC} > c_{DD} > c_{CD}$ and soft when: $c_{DC} > c_{CC} > c_{CD} > c_{DD}$; conversely, the Defender is of the hard type when: $d_{CD} > d_{CC} > d_{DD} > d_{DC}$ and soft when: $d_{CD} > d_{CC} > d_{DC} > d_{DD}$. Further, it is assumed that each player knows its own type, but only has a probabilistic knowledge about the other player's type. The uncertainty about the opponent's preferences is modelled as variables, where p_{Ch} is the Defender's belief that the Challenger is hard (i.e. $c_{DD} > c_{CD}$) and hence credible, and p_{Def} is that the Challenger's belief that the Defender is hard, or $d_{DD} > d_{DC}$. Conversely, the players' threats are incredible to the degree of $1\text{-}P_{ch}$ and $1\text{-}P_{Def}$, respectively (Zagare and Kilgour 2000).

2.1 Appraising the model and predicting the players

For present purposes, the key feature of the Zagare-Kilgour model is that it explores the connection between the credibility of the threat and stability of the deterrence, by classifying the players into types (hard and soft). Given what we know of pirates and navies, the game structure takes the following case-specific form: a pirate group can either initiate or not initiate (i.e. not put out to pirate, (CC)). If it believes that it is unlikely that it will run into someone who might defend, it prefers to put out to sea ($c_{DC} > c_{CC}$). Any pirate group knows that it cannot emerge as winner from a confrontation with a naval unit and would if confronted prefer to yield rather than to fight (since $c_{CD} > c_{DD}$). The choice between yielding and confronting for a pirate group is a choice between yielding and being incarcerated, or to fight, incur casualties and then suffer incarceration.[3] It may seem unlikely that a pirate boat would engage a navy frigate; however, wrongfully identifying a naval ship as a merchant

---

[3] This is due to legal constraints of the navies, which exclude arresting suspected pirates if there is no biometrical data from a previous encounter where there is indication that a member of a possible pirate crew is officially suspected for piracy.

vessel or ignorance of the naval presence could explain such forms of aggressive behaviour. Hence, a pirate group will most of the time rather yield than fight and is viewed as soft by their opponents, the naval vessels.

The navies defending, prefer to fight a pirate group (DD) rather than yielding (CD). However, this is not always an option; given the vast surfaces, a navy would have to control in order to confront every pirate group. Consequently, the pirates do not view the navies as hard until they manage to impress their credibility in countering pirate attacks. The navy-pirate interaction is not played between two distinct players, but rather by several agents, which can be categorised as distinct from one another, acting separately, but with the same objectives. This heterogeneity also pertains to the question of how much information the actors may have concerning one another—some may know more and some less within the same group at a certain time. Some naval units might have had more restricted national caveats than another units in the same task group, and some pirate groups might have been more risk-seeking than others. Both pirates and navies have ideas about the credibility of each other, i.e. both are considering which types they are facing—hard or soft. Hence, the game turns on whether the pirates believe that they will be confronted or not.

Since there is a mix of players, there will be a mix of outcomes over the time period studied and it is from these outcomes that the equilibria can be estimated. There are five equilibria in the asymmetric deterrence game, of which two are deterrence equilibria. The so-called Certain deterrence emerges when the Defender's credibility is high. When this is the case, Status quo is the only outcome of the game. By contrast, Steadfast deterrence is a deterrence equilibrium that exists when the Defender's credibility falls below the threshold for certain deterrence. Separating equilibria divide the players by their types. Under Separating equilibria, the Status quo of the game depends on the preferences of the Challenger and occurs in the intermediate range of the defender's credibility. Consequently, in the Separating equilibrium, any outcome is possible since it occurs when the defender's credibility is not sufficiently high to enforce deterrence, but not so low that the Status quo is impossible. When the Defender's credibility is low, two additional equilibria exist: Attack and Bluff. The Attack equilibrium occurs when Challenger's credibility is high and the Defender's is low, and the Bluff equilibrium can occur when both the Defender's and the Challenger's credibility is low.

The constants $\underline{d}$, $\underline{c}$ and $\overline{c}$ are threshold values (defined in the Appendix) that differentiate regions of prefect Bayesian equilibria.

By making it more likely to be detected and apprehended, the defender will increase the risk and lessen the possibility to highjack a ship successfully. The game suggests that a hard Defender always defends, and it is this belief that the naval operations need to impress the pirates with. However, the waters where the pirates operate are vast and for a navy to make its presence known may take time, and even then, the question is, to which degree this presence is felt by the pirates. Hence, the essence of the naval deterrence is how credible the pirates perceive the navies to be.

3 Research design

To consider security political concepts on a micro-level is not new in war studies or international relations theory. In her article "In Data We Trust", Kristine Eck points out

that the trend the last 10 years have been to increasingly focus on micro-level statistics for conflict analysis (Eck 2012). Deterrence is a vast research field, from game theoretical analysis (such as the model used in this study), empirical studies (see for instance the well-known studies by Huth and Russett 1984, and Lebow et al. 1990) to case-specific studies on deterrence (see for instance Morgan 2003). However, in all of these subfields focus has been on interstate interaction and little attention has been given to understand deterrence in micro-level conflicts, such that might occur in a military operation. Hence, when moving beyond the deterrence model to make use of the micro-level statistics there is a methodological gap to bridge. A different sort of study from 2010 authored by Stephen Quackenbush tests the predictive strength of the asymmetric deterrence model (Quackenbush 2010). By letting the outcomes from the game take the value of the dependent variable, Quackenbush tests if the equilibria suggested by the model are corroborated by the statistical data (Quackenbush 2010, 70). The results vindicate the model and suggest a methodological starting point for exploring deterrence on a micro- or tactical level. Since there are four possible outcomes suggested by the model, the statistic method used is multinomial logit.

3.1 Dependent variable

The study codes outcomes and then interprets the results from an equilibria point of view. Asymmetric deterrence or perfect unilateral deterrence, as it is also known have four outcomes: status quo, Defender concedes, Challenger defeated and Conflict. These outcomes are coded 1–4, respectively, on a monthly basis by comparing the number of pirate attacks with the number of disruptions carried out by the naval task forces. If there are more attacks than disruptions, this indicates that the naval task groups are not equal to their task and is therefore coded as Defender concedes *2*. If, on the other hand, there are more disruptions than pirate attacks, then the outcome is coded as Challenger defeated *3*, if the amount of pirate attacks is equal to the number of naval disruptions, then the outcome is coded as Conflict *4*. Finally, if there are no pirate attacks and no disruptions the outcome is coded as status quo *1*. To avoid that the data is skewed to outcome 2 and 3, outcome 1 and 4 are approximated, i.e. equal to or smaller than 3.

$$
\begin{aligned}
outcome\ (1) &= |Ap| + |Dn| \leq 3, \\
outcome\ (2) &= Ap + Dn > 3 \wedge |Ap - Dn| > 3 \wedge Ap > Dn, \\
outcome\ (3) &= Ap + Dn > 3 \wedge |Ap - Dn| > 3 \wedge Ap < Dn, \\
outcome\ (4) &= |Ap - Dn| \leq 3 \wedge Ap + Dn > 3.
\end{aligned}
\tag{1}
$$

where

Ap pirate attacks, and
Dn naval disruptions.

In this manner, each month in the two areas of operation takes on a value between *1* and *4*. As can be seen from the model (Table 1), under deterrence equilibria (both Certain and Steadfast), Status quo is the only possible outcome. Under Separating equilibrium, all outcomes are possible, and under Attack and Bluff equilibria all outcomes except Status quo are possible. Hence, in order to differentiate between

Table 1 Equilibrium and outcomes for the asymmetric deterrence game

Conditions	Equilibria	Outcome
$p_{Def} \geq \bar{c}$	Certain deterrence	Status quo
$p_{Def} < \bar{c}$	Steadfast deterrence	
$c \leq p_{Def} \leq \bar{c}$	Separating equilibrium	Status quo, Defender concedes, Challenger defeated or Conflict
$p_{Def} < c$, and $p_{Ch} \geq d$	Attack equilibrium (and bluff)	Defender concedes, Challenger defeated or Conflict
$p_{Def} < c$, and $p_{Ch} < d$		

which equilibrium that is prevailing and what the tendency is, the outcomes are studied first, then the impact of the variables. Finally, the outcomes are interpreted as equilibria.

The two main sources of data used are ICC-IMB Annual piracy statistics report and the EU NAVFOR Operation Atalanta State of Play operations sheet. In the ICC-IMB Annual reports, the time and place of each piracy incident is recorded.[4] The data is clustered by activity for each month ranging from 1 January 2008 to 31 December 2012 for the pirates and from 1 January 2009 to 31 December 2012 for the navies. The incidents are categorised as either occurring in the Gulf of Aden or in the Somali Basin/ Indian Ocean. The Gulf of Aden ranges from the Bab el Mandeb strait to the southwestern corner of Oman in the north, and roughly 400 nm east of the tip of the Horn of Africa. Since it is the pirates' willingness to pirate we are investigating, all pirate activity is merely identified as pirate incidents regardless if they amounted to a successful hijacking of a ship or an attempted attack. According to the statistical model, there is a possibility that the naval task groups make more disruptions than the pirates cause incidents. This is because the naval units can disrupt pirate vessels before they manages to encounter a ship. When this is the case, the shipping community does not report them as a pirate attack, since no civilian ship ever run into that pirate vessel.

3.2 Independent variables

3.2.1 Naval presence

Since the underlying hypothesis is that naval presence should impede pirate activity, the naval presence in each area of operation is the independent variable of main interest. Because of the Gulf of Aden's strategic importance, there has always been some naval presence and both France and the USA keep naval and military bases in Djibouti. The first real counter-piracy operation, EU NAVFOR Operation *Atalanta* or Task Force 465 (TF-465), was launched in December 2008, with the first ships on station January 2009.

---

[4] Regarding the ICC-IMB statistics, this was chosen because it records time and place of each piracy incident. Through these records, it can be verified in which area of operation the ship was sailing. The EU NAVFOR Op Atalanta State of Play and piracy statistics is also detailed, but changed its definition of what constitutes an 'attacked ship' from all attacks to only include attacks on international ships, discounting smaller or local ships, during the studied period. The ICC-IMB statistics is also corroborated with the statistics reported by the IMO for the same time. The ICC-IMB statistics is however not without its own problems; see for instance Sörenson who discusses the ICC-IMB data about under and over reporting in more detail (Sörenson, K. 2011. Wrong Hands on Deck– Combatting Piracy and Building Maritime Security, FOI-R–3228–SE, Stockholm)

Operation Atalanta was tasked with escorting WFP ships and the United Nations Support Operation to AMISOM (UNSOA), which was a logistical support operation of AMISOM. However, Atalanta's resources were primarily spent on improving security in the Gulf of Aden by establishing the international recommended transit corridor (IRTC) where commercial ships were advised to sail (see Fig. 2).

During 2009, two additional naval operations joined with EU NAVFOR Atalanta—the NATO Operation *Ocean Shield*, known as Task Force 508 (TF-508), and the Combined Maritime Forces Operation *Combined Task Force 151* (CTF-151). The EU NAVFOR Operation Atalanta State of Play operations sheet contains all three counter-piracy operations' ships with name of the ship, date and location. In addition, ships from outside navies, which did not form part of any of the operations but still liaised with them, were also reported (e.g. bilateral responses from India, China, Japan, Russia, etc.). The operations sheet accounts for the number of ships on station as well as the number of disrupted pirate skiffs in a given area of operation.

Just as in the case with Operation Atalanta, the force-flows to the different task forces were reportedly uneven, nevertheless each of the forces managed to keep at least two ships in the IRTC, averaging six frigates in the IRTC over time. During 2010, the UN-mandate was expanded and the naval operations reviewed their Rules of Engagement (RoE), allowing for more proactive operations (Sörenson and Widen 2014, 412). The same year the operation also got access to better air reconnaissance (EU NAVFOR Information Booklet 2012).

It is during 2009 that there was a real boost in naval presence, which not only brought up the number of vessel actively trying to secure the Gulf of Aden and hunt suspicious pirates, but also tried to coordinate these efforts. Focus for the operations was the Gulf of Aden and securing safe passage in the IRTC for the ships transiting in or out of the Red Sea. The first two years coordinating the task groups seems to have been complicated and it took time for the naval forces to act pre-emptively (Sörenson and Widen 2014). The first year the naval operations usually acted upon distress signals, but later with the new RoE, they also engaged suspicious skiffs. To what extent the naval presence influenced the decline in pirate activity is the independent variable. However, since other activity also could have affected the pirate activity two controls are considered in the expanded model.
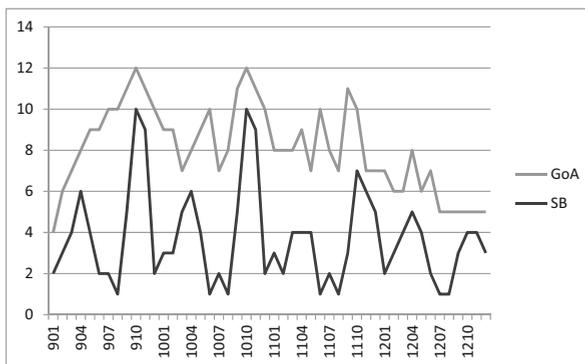
Fig. 2 Force flow of naval units in GoA and in SB. *Source: EU NAVFOR Operation Atalanta Operations Sheet 2009–2013*

3.2.2 Shipping

One important activity, which could affect piracy, is the extent the shipping community undertook security precautions. Traditionally, focus in the shipping community has centred on safety rather than security and in 2008, the shipping community's security awareness could best be characterised as uncoordinated and insufficient (Liwång et al. 2015). The increase of piracy incidents in the Gulf of Aden and off the coast of Somalia seems to have taken the shipping community off guard. It was not until 2009 that the shipping community published the so-called best management practice (BMP), instructing captains on which precautions to take before, during and after transit of the pirate waters around Somalia. The advice during this time seems also to have suffered from some confusion. While the BMP suggested that the automated identification system (AIS) should be turned off, in order to not give away a ship's position, the recently arrived naval task groups recommended that the AIS transponders should be left on, to help them to build a recognised maritime picture. Further, the 2009 edition of the BMP does not mention the naval operations as a force to liaise with (BMP 2009). During 2009, the attitude started to change within the shipping community. Although it still was uncommon to place armed guards on board, other practises such as keeping high speed (over 20 knots), hanging barbed wire along the side of the ships and water-hoses to fan off potential attackers became practises that the ships started to employ.[5]

The continuous increase in pirate attacks during 2010–2011, prompted the shipping community to rethink its security arrangements fundamentally. Many ship merchants started to offer security solutions to their clients and the IMO issued stronger policy advice regarding the Somali piracy, where the naval presence also was included as a part of a comprehensive security approach. The advice to turn off the AIS was reconsidered, the security precautions recommended by the naval task forces were implemented, and the IMO placed a liaison officer in operation Atalanta's headquarters, all in order to elevate the ships security (BMP4). The scepticism against placing armed personnel on board the ships had during this time withered among ship owners and the arming of crew members or placing armed vessel detachments teams on board became more common when transiting the Gulf of Aden and through the Somali Basin (Fig. 3).

Unfortunately, there is no detailed reporting to which extent the shipping community undertook security arrangements. However, from interviews with relevant stakeholders between 2008 and 2013, it is possible to estimate to which extent ships passing Somalia added extra security. As mentioned, from 2008 and onwards there was an increasing awareness among ship owners of the piracy threat. This increase seems to have happened exponentially; from just a small percent of the ships passing Somalia adjusting their security measures in 2008, to an estimated 20% of all ships on passage elevating their security during 2012 (interviews).[6] The data for ship passages is taken from the Suez Canal

---

[5] For a more detailed discussion on ship security, see Hans Liwång, Jonas W. Ringsberg & Martin Norsell 2013. Quantitative risk analysis – ship security analysis for effective risk control options, Safety Science, vol. 58:98–112

[6] See interviews in references.
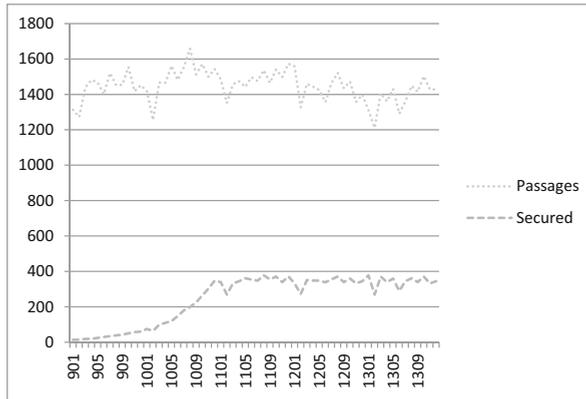
Fig. 3 Estimated proportion of ship passages with security arrangements in relation to passages over time. *Source: Suez Canal Authority: 2009–2013*

Authority which monthly report the number of passaging ships. Since the Suez Canal is the transit hub connecting the Red Sea and Indian Ocean with the Mediterranean this number, reported on a monthly basis gives a sufficient representation of the number of ships transiting the Gulf of Aden into the Indian Ocean and the Somali Basin. However, it is not the number of ships per se, which is the main interest, but the number of those ships, which undertook active counter-piracy measures.

3.2.3 Somali society

During the studied period between 2009 and 2013, the security situation in Somalia changed several times. Somalia has had a long decline in its domestic security situation dating back to the rebellion against the Said Barre regime in 1991, which propelled Somalia into civil war. The absence of security also means that reliable data on the regional stability/fragility of the Somali society is hard to come by. For instance, the reported *Gross National Product* (GDP) of Somali has only been approximated since 2006, with no official reports of GDP during 2010. For other studies on Somalia and the Somali piracy, this has been less of a problem since they have studied longer time series, where yearly fluctuations may be more relevant. Various state stability measurements such as the *Fragile State Index* (FSI) and the *IIAG Data portal*, both of which put Somalia first or second of the world least stable states, measure several factors such as demographic pressure and economic growth as well as human rights, rule of law and security. Although relevant as points of reference neither of these two indexes differentiate between the different regions in Somalia. The varying regional security situation is important for this study. Although clans often are associated with one region or another, (for instance Majerteen clan of the Darood dominates in Puntland, and Haweyi in Sothern Somalia) this seems to have primarily played a role in the beginning of the increase in piracy (see

for instance Hansen 2009 and Sörenson 2011). However, as the piracy increased it has been argued that the local security situation was determinant for the various piracy groups. For example, Kismayo became increasingly important as a pirate port, while it was not used in the beginning of the piracy during 2008 and 2009 (Sörenson and Widen 2014). There are also a correlation between the piracy attacks in Gulf of Aden, which departed from Puntland, and the piracy attacks in the Somali Basin, which made birth from southern Somalia. A rather strong indication of how turbulent the security situation was in the various regions in Somalia is the amount of fatalities, surveyed by the *Uppsala University Fatality Dataset*. From this dataset, the regional coastal areas of Puntland and southern Somalia can be extracted. Hence, the Uppsala University Fatality dataset gives a sufficient indication of the level of regional stability/instability, which also can be corroborated against the general security political trend in Somalia. The data is reported on a yearly basis and is therefore divided to fit the monthly structure of the study.

4 Results

The outcomes are studied first, and then the variables are tested in order to see which one best explains the outcomes. From the results, the equilibria suggested by the model are interpreted. In the Gulf of Aden, the outcomes show an initial steady Defender concedes outcome (Fig. 4.). The outcomes changes in May 2009 into a mixture of Defender concedes and Conflict. This result lasts to November 2011, when a new mixture of outcomes pending between Conflict, Defender concedes and Challenger defeated until the end of 2012 when the Status quo dominates. In the Somali Basin, the outcomes vary from an initial period of Status quo and Defender concedes up until August 2010 when there is a mix of all outcomes. As can be seen in Fig. 4, the outcomes move towards the Status quo in the Gulf of Aden, whereas in the Somali Basin the outcomes are scattered.
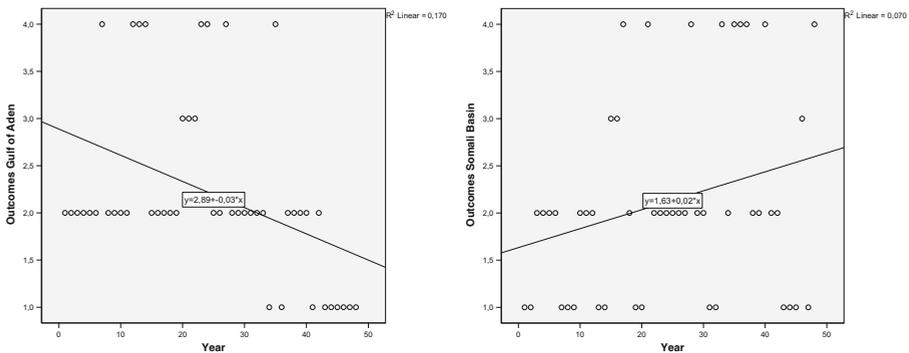
Fig. 4 Outcomes in Gulf of Aden and Somali Basin. Outcomes: 1 = status quo, 2 = defender concedes, 3 = challenger defeated, 4 = conflict; monthly index 1–48, where 1 = January 2009… 48 = December 2012

The first model tests the naval intervention to see how strong effect it had on the outcomes and the second model examines all three variables explanatory power. The first model examines the correlation between naval presence and outcomes in each area of operation (Table 2.). The naval operations in the Gulf of Aden had three different effects on the outcomes; adding one naval unit increased the likelihood for the outcome Challenger defeated relative to Status quo or, but to a lesser degree of Conflict relative to the outcome Status quo. A smaller effect is detectable in the outcome Defender concedes. Concerning the Somali Basin, the situation followed a similar pattern, but the effects were smaller. The naval operations in the Gulf of Aden mirrored the intensity, i.e. when the pirates increased their activity so did the naval task groups. In the Somali Basin, the situation is similar, but the effect by the naval task forces was smaller (compare for instance outcome Challenger defeated, where the log odds, the effect, is almost double in the Gulf of Aden compared to the Somali Basin). This corroborates that the naval units focused on the Gulf of Aden and the IRTC.

The second model (Table 3) considers the independent variable and the controls. In the Gulf of Aden, the effect of the naval operations is not considerably different after adding the control variables. The variable Ship security is non-significant in the Gulf of Aden. The control Regional fatalities is strong in the Challenger defeated outcome. The Challenger defeated outcome is an important outcome since it is in this outcome pirate incidents are trumped by naval disruptions. These outcomes also occurred during 2010 during a turbulent period in Somalia. Studying the actual number of piracy attacks during the months with the Challenger defeated outcome suggests that the piracy in the Gulf of Aden was decreasing since the number of attacks were 4–6 per month. This can be compared with 10–22 pirate incidents the months before. Hence, the naval presence and the turbulence in Puntland seem to have had an impact on the piracy.

In the Somali Basin, the effect of the naval intervention is amplified when adding the controls. The naval intervention is also what best explains the outcomes. The Ship security control is non-significant and only with a marginal effect on the outcomes. In addition, the regional fatalities were non-significant. It is important to remember that although the control variables only display some effects and often are non-significant in the model; this does not mean that they did not impact the piracy. What the model show is that from deterrence point of view the naval engagement is what best explains the

Table 2 Multinomial Logit Results without Control Variables

Outcomes[a]	Gulf of Aden			Somali Basin		
	B	Sig.	Exp (B)	B	Sig.	Exp (B)
Defender concedes						
Navy	0.895	0.007**	2.448	0.747	0.007**	2.11
Challenger defeated						
Navy	1.668	0.003**	5.303	0.838	0.017*	2.312
Conflict						
Navy	1223	0.002**	3.398	0.687	0.021*	1.988

$*p < 0.05$, $**p < 0.01$

[a] Reference category status quo

Table 3 Multinomial logit results with control variables

| Outcomes[a] | B | Gulf of Aden | | | Somali Basin | |
		Sig.	Exp (B)	B	Sig.	Exp (B)
Defender concedes						
	Navy	0.881	0.1	2.414	0.816	0.009**	2.262
	Ship security	−0.159	0.054	0.853	0.027	0.082	1.008
	Fatalities	0.543	0.156	1.72	0.077	0.137	1.022
Challenger defeated						
	Navy	1.719	0.030*	5577	1.65	0.033*	5.205
	Ship security	−0.156	0.061	0.856	−0.065	0.226	0.934
	Fatalities	4.833	0.00**	125,528	−0,273	0.21	0.799
Conflict						
	Navy	1.171	0.046*	3226	1042	0.013*	2.327
	Ship security	−0,156	0.06	0.856	0.012	0.032**	1.012
	Fatalities	0.74	0.06	2095	0.013	0.412	1.013

*$p < 0.05$, **$p < 0.01$

[a] Reference category status quo

outcomes in most of the cases. However, the outcomes also should go towards a Status quo outcome in order to make the case that the naval presence also led to the deterrence of the Somali piracy. Hence, the equilibria must be interpreted from the outcomes.

5 Equilibria interpretations

The results from the previous section showed that in the Gulf of Aden the outcomes went from a mix of Defender concedes and Conflict during 2009 to a Status quo outcome in the end of the studied period. These outcomes also correlated strongly with the naval interventions. Given the Zagare-Kilgour model for asymmetric deterrence, it seems plausible that the various pirate groups did not perceive the naval forces as a threat in the beginning of 2009, and an Attack equilibrium prevailed in the Gulf of Aden (see Table 4). The equilibrium occurs when the Defender's credibility is low ($p_{Def} < \underline{c}$ and $p_{Ch} \geq \underline{d}$). In this equilibrium, the Challenger always initiates. This can be interpreted as that the pirates were unaware or disregarded the news about naval

Table 4 Equilibria interpretation of the Gulf of Aden Jan 2009–Jan 2013

Time	Outcomes	Equilibria
Period 1 (1 Jan 2009–1 Oct 2010)	Defender concedes, Conflict	Attack equilibrium
Period 2 (1 Oct 2010–Jul 2012)	Defender concedes, conflict, challenger defeated	Separating equilibrium
Period 3 (Jul 2012–Jan 2013)	Status Quo	Steadfast deterrence equilibrium

presence, i.e. viewed the Defender as soft. Violent confrontation did also occur during this period. On 29 March 2009, the German naval tanker FGS Spessart was attacked off the coast of Somalia by pirates, probably mistaking it to be a cargo ship. The pirates were pursued and taken into custody by the German Navy (Guardian 30 March 2009). The incident suggests that the pirates had not updated their information about the risks with pirating, misjudging the Defender's credibility.

When the pirates realised that the threat had been underestimated, the Gulf of Aden area was disfavoured. However, the pirates' ambition seems not to have changed since the amount of attacks continued and increased during 2010 and 2011—but not in the Gulf of Aden (see Appendix). Differently put during 2009, the pirates updated their belief in the Defender's credibility, revised it and as a consequence the Attack equilibrium in the Gulf of Aden changed to a Separating equilibrium during 2010, and later to a Steadfast deterrence equilibria during 2012. The division by type is interesting since it suggests that, while some groups, perhaps with bad information stayed in the Gulf of Aden, others changed to the Somali Basin.

The results also show that in the Somali Basin the naval operations intensity correlate with the pirate activity. However, unlike the situation in the Gulf of Aden, there is no trend suggesting that pirate groups viewed the threat posed by the navies as a serious impediment. In the Somali Basin, the naval forces' credibility seems to have remained low, indicating that the pirate groups viewed the forces as soft in the sense that it was unlikely to run into them, although they still conceded when confronted by a naval unit. Given that the outcomes continued to be mixed throughout the duration of the studied period; a fair interpretation is that the Attack equilibrium remained until 2011 (Table 5.).

From 2011, it is possible to interpret the outcomes in terms of a Separating equilibrium; however, given the intensity of piracy in the Somali Basin during 2011, it appears more plausible that this period should also be classified as an Attack equilibrium. A crux is that according to the definitions the outcome Status quo is excluded as a possible outcome in the Attack equilibrium, and during this period there are instances of Status quo outcomes. However, when examining these periods, they all coincide with the monsoon seasons. If the periods are compared with the Gulf of Aden, where the monsoon does not seem to impede piracy in the same way, the most plausible interpretation is that they are weather dependent in the Somali Basin and that they had little to do with the general willingness to pirate. Consequently, if one accepts the explanation that these periods are weather dependent, then it is still reasonable to conclude that the attack equilibrium dominated the Somali Basin throughout the studied period.

A naval deterrent hinges on the credibility the naval task force manages project. When the naval units during the counter-piracy operations managed to make its presence known, it also appeared as more credible in its commitment to the Challenger. With increasing credibility, the challengers, in this case the pirates, started to avoid

Table 5 Equilibria interpretation of the Somali Basin Jan 2009–Jan 2013

Time	Outcomes	Equilibria
Period 1 (1 Jan 2009–1 Jan 2013)	Status quo, Defender concedes, Conflict, Challenger defeated	Attack equilibrium

confrontation, i.e. staying ashore, backing down or shifting to another area to operate in. This confirms that that credibility relates to a clear presence, which signals commitment. That the naval intervention managed to build credibility in the Gulf of Aden, but not in the Somali basin was probably due to that the Gulf of Aden was given a clear priority for patrolling and that the waters are narrower compared to the Somali Basin. These two factors allowed for a higher density of force presence which effectively denied the pirates access and increased the probability for the pirates to be captured (see for instance Geoffrey Till's discussion on sea-denial (Till 2013, 151). The critical mass in number of warships to accomplish a similar deterrent effect in the Somali Basin seems never to have materialised.

This conclusion should give pause for the naval strategist how to best prioritise patrols and what directives to give to the shipping community. A visible force, which detects events and always intervenes, will appear powerful and all-knowing. When such a force also is in constant communication with local communities, such as fishermen and traders and get its message across other media (e.g. in this particular case BBC Somalia), it will appear all present. The point is not that these measures were not taken by the naval task forces, because they were; it is that such efforts should have been made earlier, more emphatically and in coordination with the actual patrolling in order to create the credibility needed.

The study has focused on investigating the naval deterrent and not the decline of the Somali piracy in general. The two controls, (ii) shipping's own safety precautions and (iii) the domestic security situation have a low explanatory power for the outcomes. While it makes sense to conclude that, the naval operation did deter the Somali pirates in the Gulf of Aden and that they did not deter in the Somali Basin, this is not equivalent to explaining the general decline in Somali piracy. For such an explanation, additional factors must be taken into account and other methods and models must be used. For example the Jablonski-Oliver study of 2013, mentioned in the beginning of this paper, shows convincingly the importance of economic opportunity and its effects on piracy (Jablonski and Oliver 2013). Similarly, the Shortland-Verese study point out that the extent to which the local communities were willing to protect the various pirate groups also was important for understanding the mechanisms necessary for the Somali piracy (Shortland and Varese 2014). What this study has shown is that the naval forces were central to deterring the pirates from operating in the Gulf of Aden, but probably insufficient to have a decisive deterrent effect on the piracy in the Somali Basin.

6 Conclusions

This article has appraised the deterrence posed by the naval units patrolling the Gulf of Aden and the Somali Basin during the period 2009–2013. By relying on a model for asymmetric deterrence the outcomes of the navy-pirate interaction has been studied for the two areas of operation, the Gulf of Aden and the Somali Basin. While it can be shown that the navies impacted the Somali piracy, the trends for the two areas of operations differ. When categorising the outcomes in terms of equilibria the trend in the Gulf of Aden is that the equilibria went from an Attack equilibrium via a Separating equilibrium to a Steadfast deterrence equilibrium. In the Somali Basin, the most plausible equilibria interpretation is that an Attack equilibrium prevailed throughout the studied period. The reason for the

differences between the two areas of operations is likely due to that the naval operations prioritised the Gulf of Aden; and since this area is also narrower, the density of ships effectively denied the Somali pirates access, thus deterring them from operating. In the Somali Basin, the navies never managed to deter the Somali pirates. The failed deterrence in the Somali Basin was probably due to the long southern coast of Somalia and the vast area of operation, which would have required an amount of vessels that simply were not at the operations' disposal. Hence, the study shows how important it is for a navy tasked to deter an opponent to communicate a credible threat in order to effectively deter. This credibility will only appear when the threat is communicated to the opponent. Hence, for future naval engagement how to prioritise force presence and communicate resolve will be critical. In such a context, other media than actual force presence may help to communicate a task force's credibility. While the results from this study are quite clear, they are also case specific, and hence explaining the general decline in Somali piracy beyond the naval engagement will have to involve other types of studies with different methods and different data.

The study has used micro-level data to appraise a deterrence situation on tactical level. By merging two research traditions, deterrence theory and conflict analysis, the tactical situation has been analysed with relatively clear results. Analysis of this kind might also in the future further our understanding of military operations, their scope and their limits.

Appendix

Asymmetric Deterrence Game

The asymmetric deterrence game divides the players by type. The uncertainty is modelled as binary variables, where p_{Ch} is Defender's belief that Challenger is hard (i.e. $c_{DD} > c_{CD}$) and hence credible, and p_{Def} is Challenger's belief that Defender is hard, or $d_{DD} > d_{DC}$. Conversely, the players' threats are incredible to the degree of $1 - P_{ch}$ and $1 - P_{Def}$, respectively, or $c_{CD} > c_{DD}$ for Challenger and d_{DC} and d_{DD} for Defender. The thresholds are defined as follows:

$$c_- = \frac{c_{DC} - c_{CC}}{c_{DC} - c_{CD}}, \quad \bar{c} = \frac{c_{DC} - c_{CC}}{c_{DC} - c_{DD+}}, \quad d_- = \frac{d_{CC} - d_{DC}}{d_{CC} - d_{DD-}}$$
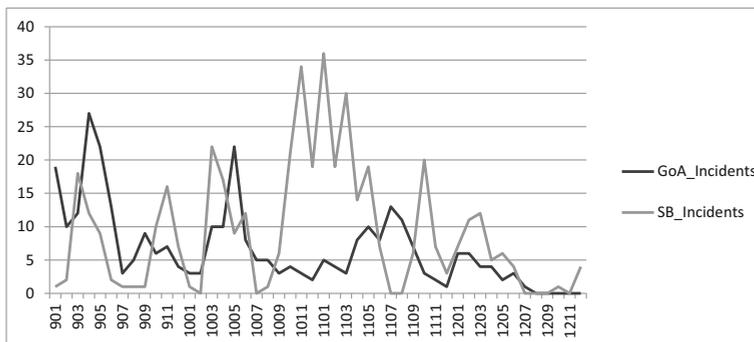
Fig. 5 Pirate attacks in Gulf of Aden and Somali Basin Source: ICC-IMB Annual piracy statistics report 2009–2013

The equilibria are represented by probability combinations expressed as [x_H, x_s; y_H, y_S], where:

x_H the probability that a Hard Challenger choses D at node 1
x_S the probability that a Soft Challenger choses D at node 1
$y_{H,}$ the probability that a Hard Defender choses D at node 2
$y_{S,}$ the probability that a Soft Defender choses D at node 2

The equilibria are defined as: Certain Deterrence exists when [$x_H = x_s$, $= 0$; $y_H = 1$, $y_S = *$] and $p_{Def} \geq \bar{c}$, and Steadfast Deterrence when [$x_H = x_S = 0$; $y_H = 1$, $y_S = u$] and $p_{Def} < \bar{c}$. Separating equilibrium exists when [$x_H = 1$, $x_S = 0$; $y_H = 1$, $y_S = 0$] and $c \leq p_{Def} \leq \bar{c}$. Bluff equilibria exists when [$x_H = 1$, $x_S = v$; $y_H = 1$, $y_S = u$], $p_{Def} < c$, and $p_{Ch} < d$. Attack equilibrium exists when [$x_H = 1$, $x_S = 1$; $y_H = 1$, $y_S = 0$], $p_{Def} < c$, and $p_{Ch} \geq d$.

References

Eck K (2012) In data we trust. Conflict Coop 47(1):124–141

Hansen SJ (2009) Piracy in the greater Gulf of Aden. Norwegian Institute for Urban Regional Research, NIBR 2009:29

Huth P, Russett B (1984) What makes deterrence work? Cases from 1900 to 1980. World Politics 36(4):496–526

Jablonski RS, Oliver S (2013) Political economy of plunder: economic opportunity and modern piracy. J Confl Resolut 57(4):682–708

Lebow R, Stein G, Janice (1990) When Does Deterrence Succeed and how do we know? Canadian Institute for International Peace and Security, Ottawa

Liwång H, Ringsberg JW, Norsell M (2013) Quantitative risk analysis – ship security analysis for effective risk control options. Saf Sci 58:98–112

Liwång H, Sörenson K and Österman C (2015) Ship security challenges in high-risk areas: manageable or insurmountable? WMU J Marit Aff 14(3)

Menkhaus K (2007) Governance without government in Somalia: spoiler, state building and the politics of coping. Int Secur 31(3):74–106

Morgan P (2003) Deterrence now. Cambridge University Press, Cambridge

Quackenbush SL (2010) General deterrence and international conflict: testing perfect deterrence theory. Int Interact 36:60–85

Rasmusen E (2007) Games and information, 4th edn. Blackwell Publishing, Oxford

Shortland A, Percy S (2013) Governance, naval intervention and piracy in Somalia. Peace Econ Peace Sci Public Policy 19(2):275–283

Shortland A, Varese F (2014) The Protector's choice–an application of protection theory to Somali piracy. Br J Criminol 54(5):741–764

Sörenson K (2011) Wrong Hands on Deck–Combatting Piracy and Building Maritime Security. FOI-R–3228–SE, Stockholm

Sörenson K, Widen JJ (2014) Irregular warfare and tactical change–the case of Somali piracy. Terrorism Political Violence 26(3):399–418

Zagare F, Kilgour M (1993) Asymmetric deterrence. Int Stud Q 37(1):1–27

Till G (2013) Seapower - A Guide for the Twenty-First Century, 3rd edition. Routledge

Zagare F, Kilgour M (2000) Perfect deterrence. Cambridge University Press, Cambridge

Data

IMB-ICC Quarterly and Annual Piracy Report 1992–2013

EU NAVFOR Operation Atalanta Operations Sheet 2009–2013

Suez Canal Authority, Passages 2009–2013

Uppsala University Fatalities dataset; Puntland and Southern Somalia 2009–2013

Reports, New-Articles and other Documentation

Best Management Practice , International Maritime Organisation, February 2009
Best Management Practice 4, International Maritime Organisation, March 2011
EU NAVFOR. 2012. Information booklet
The Guardian 30 March 2009 Blundering Somali Pirates arrested after attack on anti-pirate ship. http://www.guardian.co.uk/world/2009/mar/30/pirates-attack-nato-ship. Accessed 9 Jan 2016
UNSCR 19 July 2008. "Report of the Secretary General on the Situation in Somalia" S/2008/466
World Bank Report. 2013 The pirates of Somalia – ending the threat and rebuilding a nation. Washington DC

Interviews

Wallenius Shipping, Stockholm December 2008
World Food Programme, Nairobi February 2009;
HSweMS Malmö, Djibouti April 2009;
Seychelles Port Authority, Victoria June 2010;
Force Headquarter Operation Atalanta HSweMS Carlskrona, Gulf of Aden July 2010;
European Commission Somalia Desk, Brussels November 2010;
Gulf Agency Company, Stockholm September 2011;
Operational Headquarter Operation Atalanta, Northwood November 2010 and April 2013