Deterrence Models and their Targets

This paper compares and discusses the Classic Deterrence theory’s model with its proposed replacement the Perfect Deterrence theory’s models and their respective fit to intended targets. Since the idealized model in this context can be problematic the paper studies the deterrence models’ foundations from five criteria proposed to evaluate idealized models. It is concluded that since the PDT models has a whole system of targets while CDT only relate to one specific target comparison will be case dependent. Although the PDT models give a more complete account of deterrence dynamics, as a consequence they lose simplicity. Conversely, the simplicity of the CDT model which in essence has been its strength is also its weakness in inexplicable phenomena. Consequently, it will depend on the type of inquiry, which model should be put to use. The relative nature of models does not preclude the fact that PDT models have enriched the research field by unravelling possible dynamics and in so doing extended the research in many related fields.

*Karl Sörenson*

\*Swedish National Defence College and Royal Institute of Technology

**Corresponding author:**

Karl Sörenson, Swedish National Defence College, Drottning Kristinas väg 37/ Box 27805/115 93 Stockholm, Sweden

Email: karl.sorenson@fhs.se
Introduction

In the discussions on deterrence and crisis-bargaining new models are often presented and new games suggested. Lately new models have emerged which differ from previous ones with regards to assumptions about rationality, credibility and information. While many models were developed during the cold war and consequently model behavior from that era (e.g. Herman Kahn, Thomas Schelling etc.), new models stretch the scope to also include asymmetric deterrence and various forms of crisis-bargaining (e.g. Frank Zagare, Mark Kilgour and Branislav Slantchev, etc).

In 2000, Frank Zagare and Marc Kilgour published what to date might be their magnum opus in a series of research articles on deterrence entitled Perfect Deterrence. The work merges several of Zagare’s and Kilgour’s earlier work into one cohesive body. Since then the Perfect Deterrence Theory (PDT) of Zagare and Kilgour has been discussed and expanded upon by various scholars Branislav Slantchev, for example, relies on a part of the PDT when he considers credibility in connection to crisis-bargaining and costly signaling (Slantchev). From a theoretical point of view, what is essential when compared with Classical Deterrence Theory (CDT) is PDT’s focus on, credibility and information. But it is not only the models theoretical aspects that seem to have drawn attention to PDT. When Stephen Quackenbush investigates the PDT validity through empirically testing he concluded that PDT performs well (Quackenbush 2010). What is particularly interesting with the Quackenbush investigation is that it possibly bridges a gap between the empirical investigations of deterrence and the formal theorizing of deterrence, two fields which have co-existed, but rarely overlapped.

The advantages of PDT over other understandings of deterrence are something which Quackenbush himself investigates further when he compares PDT with its predecessor CDT (Quackenbush 2011). Quackenbush restates and reduces some of the fundamental underpinnings of the Classical Deterrence Theory (CDT) and then compares and contrasts them with those of Perfect Deterrence Theory (PDT). Quackenbush’s analysis focus on the theoretical foundations of CDT and PDT such as credibility and bargaining, but also various concepts associated with deterrence, such as mutual vs. unilateral deterrence, and general vs. immediate deterrence. Furthermore Quackenbush also considers empirical tests of deterrence. The same problems of CDT are also treated by Zagare and Kilgour themselves as reasons for developing their models and PDT (Zagare & Kilguor 2000). Quackenbush’ major criticism of CDT, which also is in line with the Zagare and Kilgour critique, is that its concept of credibility is faulty (which in turn has to do with its understanding of rationality). He then continues to point out that the crucial difference regarding rationality and credibility presented in PDT is suspiciously missing from contemporary work such as Patrick Morgan’s Deterrence Now (Quackenbush 2011).

Hence, one of Quackenbush’ underlying conclusions seems to be that Morgan and perhaps other scholars have not fully understood or appreciated the benefits of the PDT over CDT. Quackenbush
himself concludes however that PDT should replace CDT “as the most appropriate basis for further theoretical development, empirical testing and application” and in so doing suggesting something along the lines of a change in paradigm from CDT to PDT (Quackenbush 2011, 762).

Although the Quackenbush and Slantchev seem to consider PDT as an improvement, the PDT has also had its detractors. Fred Lawson criticizes PDT to be too rationalistic in its approach and seem to favor the CDT as a better foundation for policy analysis. At the heart of Lawson’s critique is the mathematics of PDT which becomes in his view too complex to bring clarity to important questions. When summing up the views on PDT it seems like its status is somewhat ambiguous, some seem to consider it as an improvement like, Quackenbush and Slantchev, some, like Lawson, view it as a dead-end, while others such as Morgan simply ignores it.

While some of the varying views perhaps can be found in the personal preferences for certain types of theories, one aspect has to do with how models should be viewed in science and what ultimately represent. Both Quackenbush and Lawson are on the one hand appraising the theories, but on the other hand they are considering the significance of the model, which in both theoretical positions are crucial. With few exceptions the model itself is rarely put front and center in a scientific discussion. This is unfortunate since Lawson as well as Quackenbush are both under the misapprehension about what a model ever could be. A similar example is the discussion that took place in 1999 between model critics and model proponents. In essence, Lawson’s critique of PDT is reminiscent of the position taken by Stephen Walt in his article “Rigor or Rigor Mortis” which aimed at the simplifications, miscomprehensions and unwarranted mathematical complexities which he saw as a consequence of the influx in models use in security studies (Walt 1999). In a reply Robert Powell defended the increasing use models as he saw as bringing clarity and helped with predictions (Powell 1999). Although Walt’s position is questioning the model’s very status, Powell does try to show that even a simple model can yield powerful insights he does not develop this line of the argument further. He never gets so far that he really grapples with what a model actually is and how it can be appraised. It is this very gap of how we view models that this article hopes to clarify, specifically with regard to the deterrence models.

The article ascertain what the CDT and PDT models could be understood to be from a model perspective and in so doing it also investigate what their respective scopes and limits are. For consistency in this investigation we ask three questions: (i) What type of models are the deterrence models? (ii) Are there a substantial difference between the Classic Deterrence models for deterrence and the Perfect Deterrence model? (iii) How can we appraise these models, given their type and intended targets?

In order to answer question one (i) I try by means of model theoretical practises of categorisation say something about the models’ type and their intended target systems. I then turn to the second (ii)
question and consider strengths and weaknesses of the different model types, and the models relationship to their targets. In connection to this I also ponder flaws, problems and paradoxes and potential unexplained phenomena. Finally, I try to answer the third question (iii) and say something about what the models gives us from an epistemological point of view.

The article follows a path where the two models are generally explained and we then move to consider the model theoretical specifics of them in order to finally conclude what the inquiry has given us. The two main conclusions are that 1) a model can only be appraised in relation to its intended target or target-system and from this perspective it is in a sense meaningless to compare them since the PDT models has a whole system of targets while CDT only relate to one target. Further it is concluded 2) that is gained in model completeness by PDT is also lost in simplicity, conversely the simplicity of the CDT model which in essence has been its strength is also its weakness in inexplicable phenomena. From this point of view it will always depend on the inquiry one aims to investigate which model should be used. Although it is concluded that the status of a model is more relative then first might appear it does not follow from this that the creation of the PDT models have not enriched the field by unravelling possible dynamics and in so doing furthering research in many related fields.

**The Idealization of Conflict**

A potential problem when investigating models is whether a theory can be equated with its models. Put in the form of a question: Does critique of a theory have implication for its associated model(s) or vice versa? From a scientific point of view the connection between model and theory is complex. On the one hand models are often associated with a theory and developed in order to explain or predict something in association to a theory. On the other hand, a theory might be recognised as overplayed, while a model developed within the context of a theory may still survive and be relevant. And just as a model which represent poorly does not automatically invalidate a theory, a theory which is dead does not take the models with it to its grave. Hence, the understanding of the relationship between theory and model must always be made case specific.

In the case of PDT and CDT the à priori answer is twofold: On the one hand, CDT is not just one deterrence model because the model contributions have been many in the course of CDT history. There is no one model which fully accounts for all the added and subtracted functions which have been tinkered with throughout the course of the CDT debates. Hence, some care is needed so that refuted aspects of CDT models are measured in relation to how central they are to the CDT in general – marginal problems can only warrant marginal critique. A similar point is made by Patrick Morgan who thinks that although there may be many differing contributions to deterrence theory they all belong to the same research field and theory (Morgan 2003). Zagare and Kilgour themselves seems to be in agreement with this account, but they separate the so-called structural deterrence theorists from
the decision theoretic theorists. Since this paper will focus on the modelling side of CDT, the structural theorists can be omitted.

The PDT can on the other hand to a larger extent be equated with its models. This is because PDT does not conclude anything which is beyond the claims of its models. Zagare and Kilgour’s work centre around their deterrence models and coda, and nothing beyond the models are asserted. Further, and as mentioned earlier, this is not a matching comparison, it is an investigation to what extent new phenomena are unveiled or different phenomena uncovered by the PDT compared to the CDT models. What this relationship is will be something that I return to when considering how we can understand the relationship between model and theory. Hence, while the analysis of PDT can be made quite specific, it must be remembered that the analysis of CDT can only be approximate. Nevertheless, there are some important differences between the two even when one is distinct and one is an approximation.

**The classical deterrence model** was modelled after the Cold War nuclear threats. In modelling this conflict game theorists assumed two players, or states, who both want to deter the other from gaining a strategic advantage, but at the same time trying to attain an advantage for them self, while making sure full scale conflict is avoided since it is equated with nuclear war. This was because the general situation was considered to be like this: both the US (and some if its allies) and the Soviet Union retained nuclear weapons which at a moment’s notice could be launched. It was in the interest of both not to be taken off guard, because then the opponent could launch and gain a strategic advantage. To make sure this advantage would not be too advantageous both powers also possessed the capability of retaliatory launches, i.e. if one was to launch and this attack could promptly be revenged by the surprised party, ensuring destruction in a similar way. This strategic situation became known as MAD or mutually assured destruction. It was believed with sufficient nuclear arsenals and MAD in place the bipolar world would be stable, since execution of a threat would mean destruction of oneself. Many of the most prominent scholars such as Brodie, Kahn, Schelling, etc., considered various aspects of this, in many ways strange, strategic dilemma. In the game, in which most of the academics agreed with, the game takes the classical form of chicken-race. Each player has two choices, she can either cooperate (C) or defect (D). If both states cooperate the Status Quo relationship will remain intact (CC). If one player choses to cooperate but the other choses to defect the ‘defector’ gain 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 CD, CC, DC, DD and for B is DC, CC, CD, DD (since 4>3>2>1). If A decides to initiate we call her Challenger and B for Defender and vice versa. 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 Nash equilibria DC and CD, depending on who moves first.
As can be seen from the strategic form above (fig. 1) the assumption is that conflict (DD) is the worst outcome. Zagare, Kilgoure, Quackenbush and Slantchev all critique this assumption for being too strong. If conflict is to be the worst outcome for all parties, which in itself is not an unreasonable thing to model given that conflict meant the execution of nuclear threats, then it also imply that the game is not stable. The incentives to make a first move are stronger than both status quo and conflict. For Defender this means that she should concede since DC > DD, but if this is common knowledge than Challenger will always initiate, and Defender will always concede. Since the game is symmetric this would imply that the structure is instable, i.e. it would be rational to challenge, since CD > DD, and irrational to defend. This fact indicate that there is an ill fit between the model and target since the Cold War relationship normally was described as stable and although crises broke out, they never amounted to full out conflict. The extensive-form of the game is presented in figure 2.

This problem is well known and there were principally two ways of dealing with the model’s shortcoming in terms of capturing the stability of the Cold War. One way was to make Defender make binding commitments to a defence strategy, thus eliminating concession as an option for the defender (Powell, et al). Many theorists have tinkered with this solution, but given that the game structure is
kept intact, Defender must behave irrationally, i.e. choosing DD, although DD<CD. Since the move is not sub-game perfect it violates the model, indicating an ill fit once more.

The second way to account for the stability is to involve threats with risk-taking (Schelling, Kahn etc). By threatening the opponent with full escalation, it will be up to Challenger to initiate, but making sure that she knows that with an escalation there is also a risk of conflict, which would be damaging to both parties. If Challenger deems the threat credible, it would not be in Challenger’s interest to initiate since there is a non-trivial risk of full out war. Thus, the second solution involves the idea of credible threats, leaving it up to the opponent to take the risk of full-fledged conflict. In this context a credible threat are equated with threats which are believable, but since it would be irrational for defender to go to conflict when concession still would yield a better outcome, the threat from defender is not believable, and hence not credible. It is this paradox that Zagare and Kilgour focus on when they formulate their critique and subsequent response.

In the Perfect Deterrence Theory model of mutual deterrence the game takes the following form: Just as before player A and B have two choices they can either cooperate or defect, there is like before four possible outcomes Status Quo, Advantage A, Advantage B and Conflict. The preference between the outcomes are however different. For A: \(a_{DC} > a_{CC} > [a_{DD} \text{ and } a_{CD}]\) and for B: \(b_{CD} > b_{CC} > [b_{DD} \text{ and } b_{DC}]\), hence the Conflict and Advantage for the opponent are left open. Further each player’s estimate of the other player’s preference of Conflict over concession is taken as measure of the other player’s credibility. The strategic form is listed below in figure 3 (Zagare and Kilgour 2000):

![Figure 3](image)

The preferences a player may or may not have determine which types of player she is: A is considered of the Hard type when A: \(a_{DC} > a_{CC} > a_{DD} \text{ and } a_{CD}\) 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: \(b_{CD} > b_{CC} > b_{DC} \text{ and } b_{DD}\).

In order to model the uncertainty with regards to the player’s preferences the model introduces independent binary random with known distributions. 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 2.1. However, the only time when deterrence succeeds is in the case when the players are credible, i.e. Hard, 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 with credible threats. However, it is not always the case, perhaps even rarely, that the game exists under complete information criteria. The incomplete information of the game takes a rather different form. With incomplete information backwards induction is not possible; hence, the game cannot be solved in this way. Under incomplete information there are 8 possible equilibria, which can be categorized in three groups, depending on 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). Also what is important is that the mutual game can be turned into a game of chicken, but it does not have too. 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 is being played.

In their Perfect Unilateral Deterrence Zagare and Kilgour explore a game where: “Deterrence of a challenger by a defender is modelled by explicitly relating uncertainty and credibility of retaliatory threats to the stability of an asymmetric deterrence relationship” (Zagare and Kilgour 1993). In this game the agent’s or player’s role are categorized into a Challenger (C) and a Defender (D). The Challenger decides whether to initiate or to not initiate a challenge. If the Challenger decide for the second alternative the game ends in a status quo outcome (or CC in Fig. 1). If, on the other hand, the Challenger decides to challenge the Defender, the Defender must decide whether to allow it or to resist it. If Defender chooses the former alternative, the game ends and the Challenger gain an advantage. If, however, the Defender decides to meet the challenge, the Challenger must decide either to back down and give Defender an advantage (CD), or to face confrontation and a possible defeat (DD). The extensive form of the game is presented in Figure 1.
Zagare and Kilgour place some restriction on the game to make it realistic. Their first restriction is that Challenger strictly prefers an advantage to the Status Quo (CC). The Defender could however prefer either to stay in the status quo or face down a challenge. Secondly, each player prefers either winning without a confrontation, CD for Defender and DC for Challenger, or the Status Quo (CC), to conflict, DD. Finally, each player knows the other’s preferences satisfy these restrictions, but is uncertain about whether the opponent prefers conflict, DD, to capitulation, CD for Challenger and DC for Defender. The preferences a player may or may not have decide which type of player she is Challenger considered of the Hard type when: \[c_{DC} > c_{CC} > c_{DD} > c_{CD}\] and Soft when: \[c_{CD} > c_{CC} > c_{CD} > c_{DD}\]. Conversely Defender considered Hard when: \[d_{CD} > d_{CC} > d_{DD} > d_{DC}\] and Soft when: \[d_{CD} > d_{CC} > d_{DC} > d_{DD}\]. Zagare and Kilgour further assume that each player know her own type, but only has a probabilistic knowledge about the other’s type. The uncertainty with regard to the players’ preferences is modelled as independent 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 (Zagare & Kilgoure 2000).

In the Perfect Unilateral deterrence game with incomplete information there are five equilibria: two Deterrence equilibria (called Certain Deterrence and Steadfast Deterrence), one Separating equilibrium, one Bluff equilibrium and one Attack equilibrium. Hence, since the roles are distinct in asymmetric deterrence there is only one equilibrium where the Challenger prevails.

[My excuses for not explicating the Perfect Extended Deterrence model and problematizing PDT in general, which was meant to be done here – about a page.]

**What makes a model of S? - What is being modelled**

The implicit assumption among philosophers is that for a model to truly represent, it must depict real targets, entities which can be found in the real world. Appraising models carries with it several difficulties such as correlation and intention (Hodges 2009), resemblance (Mäki 2009) and similarity (Giere 1988). These difficulties are compounded when comparing two models, because not only do we need to make sure of the fit between model and target (often expressed as if \(M\) is a model of \(S\)) but we will need to understand how the models differ and if their intended targets differ as well. There are no straight criteria for these sorts of comparisons, but the modelling literature does suggest certain aspects which can be indicative when appraising two models and their intended targets.
From Hodges intentional point of view there are different kinds of models; the model of the weight on a spring aims at depicting the physical dynamics of a spring, whereas a find-algorithm for a computer programme is an instruction for function. In the first case we are dealing with a descriptive model and in the latter case a functional model. Hybrids of the two model-versions exist, such as a model of a wind tunnel which is a description and an instruction (Hodges 2009). A similar view is echoed by Michael Weisberg in Simulation and Similarity, however, where Hodges considers functional models as a class in its own right. Weisberg for his part, differentiate between mathematical models, procedural models and scale models. Hence, his categories differ from Hodges with regards to scale models.

The multitude of classes of models in model theory need not concern us here, since both the classifications contain the relevant classes for our purposes; the mathematical model. However, if model $M$ is to represent a system $S$, where $M$, as in this specific case is a situation of deterrence between states, and $S$ is a situation, the model-target relation does not satisfy the standard of target-direct modelling since this is normally understood as the relation between a single target, like that of a building, and a single model, like that of a blue-print of said building. Since modelling quarrelling states necessitates a different approach the representation needs to be understood in a different way. (Morgan and Knuutila 2009, see also Weisberg 2013). First of all, when modelling several instances of the same phenomena one normally does not think of it in terms such as target-direct modelling, but rather in terms of a target-system of a phenomenon, it is an abstracted version of the intended target (Giere 1988). When models not only represent a target-system, but also put stress on a certain aspect of a target-system they are often referred to as idealised models. Models of this kind are common in the social sciences such as economics and M-representation of target-systems tends to be different in idealised models compared to non-idealised models. Here I will not diverge from the general understanding but treat the deterrence models as idealized models. The question about their intended target will remain but it will need to address other criteria than perhaps a model of a tower or a code-breaking algorithm would need to do.

**Simplicity or Completeness - How is it modelled?**
Michael Weisberg proposes five criteria for evaluating idealized models; Completeness, Simplicity, 1-Casual, MAXOUT and A/P-general. Incidentally, Weisberg also considers these criteria as ideals for the modeller (Weisberg 2013).

Completeness implies two sorts of completeness; inclusion-rules and fidelity-rules. The inclusion rule states that every property of an intended target system must be included in the model. Causal and structural properties of the phenomena must be captured by the model. The fidelity-rule, in turn, demands of the modeller that every property must be accurately represented. Opposite the
Completeness criteria is the Simplicity criteria, which instructs the modeller to include as little as possible in the model, but without infringing on the fidelity rule. Perhaps quoted to excess, but unusually apt is the Einstein quote that “everything should be made as simple as possible but not simpler”. Weisberg points out that there are two dimensions of 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, i.e. when creating a model it can be helpful to start out in a simple fashion and then to the build the model bottom up to increasingly handling complexities. The Completeness and the Simplicity criteria can be understood as diametrically opposed ideals. Is simple way of relating them to each other is to think of them as what one gains in Completeness one may lose in Simplicity and vice versa. The remaining three criteria are in a sense relatable to one of the two (Weisberg 2013).

1-casual states that only the central causes of the studied phenomenon should be included in the model. Hence, 1-casual controls the level of simplicity the model can actually have, although it also is related to this ideal since it limits higher-order causes which could make up a part of a model. Just like 1-Causal is related to simplicity, so is MAXOUT to Completeness. MAXOUT instructs the modeller to be as precise and accurate as possible in construction of the model. Hence, every property which affects the target-system should be incorporated into the model, but where Completeness turns on the representation, MAXOUT relates the level of accuracy in output of the model. Hence, for maximizing a models predictive powers Simplicity may interfere with an ambition of prediction. A/P-generality relates the number of actual targets that are represented in the model. A-generality is here the actual targets that a model applies to, whereas P-generality is the possible targets the model could capture. As in the case with Completeness and Simplicity 1-casual, MAXOUT, and A/P generality relate to each other, but perhaps not in the same more fundamental way. The need for 1-casual can potentially limit precision (MAXOUT), and striving for generality could foreseeably jeopardize the 1-Causal criteria (Weisberg 2013).

Idealized models such as the deterrence models are often understood as a certain type of idealization often referred to as Galilean idealisation. These types of models are taken to distort certain properties of a phenomenon in order to study just these (Morgan and Knuutila 2009). The Galilean idealisation is often contrasted with Aristotelean or minimalistic idealisation, which tries to reduce the representation in the model to the very minimum. The two types of idealisation, mentioned here, are also associated more strongly with some of the Weisberg criteria: the Galilean idealisation is related to the Completeness criteria and perhaps the MAXOUT criteria, whereas the minimalistic idealization is associated with the Simplicity and 1-Casual criteria (Weisberg 2013). It is however important to remember that the Weisberg criteria are not evaluative properties. That a model is very simple does not imply that it is flawed or wrong, just that it is simple. A model which is very detailed can still misrepresent. What they do tell us is what properties a certain model manifest, nothing more.
The target-system of the Classical Deterrence model is in many ways a distortion of its target system. It does not try to cover the entirety of the Cold War dynamics, but a specific aspect of it, namely when it is rational to strike, when to concede, when to initiate and when to observe the status quo. It is important that the model captures the two agents’ rationale for these various choices. The model strives for Completeness of the phenomena of the Cold War, but only as a specific instance of symmetric deterrence with high stakes.

![Diagram](image)

**Figure 3.**

When representing the relationship of the Classical Deterrence model to its target (and making the steps explicit) it is clear that the CDT model is focused on a very specific instance of deterrence. As can be seen in figure 3., the CDT accounts for the Cold War nuclear deterrence, there is rarely an interest in other aspects of deterrence and the focus is firmly on the fidelity-rules as to what to include and what to exclude in the representation. Deterrence beyond the deterrence between the East and the West is not a part of the target. In other words, it is trying to give a Completeness account of its target system, but a system which is quite limited. Hence, with respect to A-generality it is questionable how generalizable CDT deterrence really is, unless identical situation appear (e.g. Pakistan and India’s mutual deterrence). From a P-generality point of view it may still have something to give to game-of-chicken-like situations. Examples of this can also be found in the CDT literature, references are often made to situations thought to be similar to the Cold War Deterrence. A favourite example is the security political situation leading up to the First World War, but there are also others. [find another example].

The principle difficulty with the CDT model seems to revolve around the 1-Causal criteria, i.e. which first order causes are allowed? Are believable threats credible, and would this cause restraint from a challenger? Would binding commitment cause the Challenger to refrain from an attack. It is difficult to ascertain the exact strength of these causes, but it seems like the modellers never managed to free themselves from 1-causal criteria. Is the model simple? Given that the Game of Chicken is an intuitive picture of what a mutual deterrent relationship is all about, it is tempting to say that it is. However, the many contributions and subsequent mathematical dressing that took place of the model during the cold
war, the Simplicity criteria cannot be answered as straight forward as first implied. From the game that we have studied here, which in a sense is in the “standard” Game of Chicken form the game should still be understood as simple. The fact that the model is simple might have had baring on the difficulty scholars and analysts have had to free themselves from the embedded assumptions, and this in turn is what has created the need to add various aspects in order from the game to represent stability. Considering the difficulty the model had to explain the stability of the Cold War, although this is what it was thought to model, the MAXOUT criteria seems rather to have come into play to amend a model which in fact was giving a misrepresentation of its target.

The many appendixes which strive to cover for various theoretical deficiencies could perhaps be viewed as indicative of a faulty model, but it could also be viewed as a model in its own right which goes to the heart of a specific aspect of the Cold War and mutual deterrence-like situations. It is simple, but this also means that for a lot of users its intuitive results are manageable and inviting creativity.

**The Perfect Deterrence Theory** models strive for a fuller application of deterrence phenomena in general. In fact the model seems to have sprung out of a dissatisfaction of the CDT model. Just as the PDT models leave open the preference for conflict over concession; it also tries to capture the credibility of the player’s threats. But it does not stop there; it tries, with the same assumptions, to model conventional mutual deterrence, asymmetric deterrence situation as well as extended deterrence situations. As can be seen in figure 4., the criterion for completeness is clearly important since the ideas of also capturing rational play (without a paradox), credibility, and unilateral as well as extended deterrence expands the intended target into a system of targets. Further, the mathematical details are worked out in quite some detail, with the Bayesian separation model intended to isolate different types of players, hence enabling more sophisticated results as well as predictions. Consequently MAXOUT is an important part for the modellers and the price is paid in terms of its lack of Simplicity.

Figure 4.

Nevertheless, it seems like only first-order causes are allowed into the model, so from a 1-causal criterion perspective higher order causes are not allowed. The generality of the model(s) are clearly
intended to cover more than the CDT model, and since the model is also divided in several version its
generality cover more phenomena (A-generality) (see Quackenbush who uses the model for conflict
prediction). The P-generality is also quite high, and it seems like the model could be put into different
context (see Sörenson 2014, who uses the PDT unilateral deterrence model for piracy prediction).

**Appraising Deterrence**

After having scrutinized the two models, their intended targets or target systems and criteria which
seems to have been guiding the modelling, can we unequivocally state which model is the best? Is
such a question even meaningful to ask?

While the CDT model is modelled after one, albeit global, conflict and its target is uniquely outlined.
It aims to represent the most important strategic aspects of the Cold War. The Game of Chicken is the
modelled which for decades was thought to more or represent this intended target of the Cold War. It
is an idealized model since its abstraction from its own target is clear. 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; Many are the contributions which over the decades have piled up, sparking
new research fields, Zagare and Kilgour are themselves children of the CDT model and the PDT has
sprung out of certain perceived deficiencies. These deficiencies, identified as paradoxes Zagare and
Kilgour, are of course problematic when viewing the model in relation to its intended target.

The PDT models several different instances of deterrence, from mutual deterrence, conventional or
nuclear, unilateral deterrence and extended deterrence. In addition, in its system of target is also the
credibility of the players an important aspect.

There is only one instance where the two models overlap and have the same target. If any comparison
is to be made it is here it must be, for all other cases PDT is the only theory which presents models
developed for other types of deterrence scenarios. When faced with a bipolar nuclear deterrence
situation which of the two models should one turn to? But also here it depends, is it to understand the
dilemma situation of two agents confronting each other, then the CDT might be more than ample
choice. The CDT has a somewhat tarnished but proven track record and its very simplicity might help
to capture the situation one is most interested in. If however, there is a need to also understand how
this relationship may change, how credibility plays out and how important information is then, the
PDT has much more to offer.

An analogy from physics might clarify this very point. When understanding the planets relative
motion to each other Newton Mechanics of point mass describe the movements of the heavenly bodies
well. However, it is not the whole story. Consequently, when sending a satellite to explore outer space
calculation of gravity will be need to avoid that the satellite gets side tracked by the gravitational pull
of some other body. To avoid this, relativity theory calculations will be needed to also account for the gravitational fields which might have bearing on satellite’s trajectory. The simplification of point mass is allowed as long that it is serving its purpose. When the purpose changes it may be the case that point mass represent less well.

In one sense the PDT could be viewed as a form of progress in the search for our understanding of the phenomena known as deterrence. The PDT model covers several aspects of deterrence-relationships symmetric, asymmetric and extended. It also appraises level of credibility in relation to information. All of this is extensions of the CDT model. So, for the scientist looking for more sophistication or a model which cover unilateral deterrence or model extended deterrence the PDT model should perhaps be favoured.

Hence, Quackenbush is right in his assertion that the PDT is a more complete account of what deterrence, but this does mean that it is better or should replace anything. Lawson is possibly right when he asserts that CDT is a better basis for policy analysis, but this can only be the case only be the case in the instance of two models’ overlap and only if this policy is not in need of any of the more developed, complete accounts of mutual deterrence. The only thing which remains odd is Morgan’s omitting of the PDT from his book *Deterrence Now*, since the creation of PDT has meant that our understanding of deterrence has become fuller and more complete.

**Conclusion**

A model can only be appraised in relation to its intended target or target-system and from this perspective it is in a sense meaningless to compare the CDT with the PDT. This is because the PDT encompasses a system of targets while the CDT only relate to one target. Further, what the PDT gains in completeness is also its loss in simplicity, and by the same token the simplicity of the CDT model, which in many ways has been its strength, is also its weakness since it leaves many questions unanswered. From this point of view it will always depend on the question which model would serve one’s inquiry the most. Although the status of a model is more relative then first might appear it does not mean that the PDT models have not enriched the field by unravelling possible dynamics and in so doing furthering research in many related fields of deterrence and elsewhere.
References


---