

Prospects of Deterrence: Deterrence Theory, Representation and Evidence

Karl Sörenson

To cite this article: Karl Sörenson (2024) Prospects of Deterrence: Deterrence Theory, Representation and Evidence, Defence and Peace Economics, 35:2, 145-159, DOI: [10.1080/10242694.2022.2152956](https://doi.org/10.1080/10242694.2022.2152956)

To link to this article: <https://doi.org/10.1080/10242694.2022.2152956>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 19 Dec 2022.



Submit your article to this journal [↗](#)



Article views: 5538



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 3 View citing articles [↗](#)

Prospects of Deterrence: Deterrence Theory, Representation and Evidence

Karl Sörenson 

KTH - Royal Institute of Technology, FHS - Swedish National Defence University, Stockholm, Sweden

ABSTRACT

Game theoretic analysis of deterrence has been criticized for not capturing how actors realistically behave. It is alleged that prospect theoretical re-modeling provides a better foundation for a deterrence theory. The article analyzes how the strategies change when a prospect theoretical function is applied to a central deterrence game. While the probability distributions changes, it cannot alter the general dynamics. When considered together with previous research, it shows that prospect theory neither can or should replace standard assumptions when constructing a deterrence theory. However, viewed as a compliment, prospect theory expands the modeling possibilities and opens up for important new aspects.

ARTICLE HISTORY

Received 1 February 2022
Accepted 25 November 2022



KEYWORDS

Deterrence theory; game theory; prospect theory

Changing World, Changing Theory

Game theoretic analysis of deterrence has been criticized for not adjusting to how actors realistically behave. The Russian invasion of Ukraine in February 2022, may be an example of such a shortcoming in analysis. While the history of the outcome of the Russian invasion remains to be written, analysts, scholars and policymakers have expressed surprise both at the strategic mistakes committed by the Kremlin as well as of the Western resolve in response to the attack on Ukraine (Dalsjö, Jonsson and Norberg 2022). Adhering to standard deterrence analysis, it could be argued that the Russian decision to invade would have been ill advised given the realities on the ground and therefore was deemed as unlikely to happen. Further, it was argued that Western help to Ukraine should be of the prudent kind in order not to provoke Russia.¹ However, if the Russian decision to invade was made in reference to an idea that they now had the opportunity to bring Ukraine under their control, the West should have deemed the likelihood of an attack as considerably higher. Similarly, the Western caution before the outbreak of the war were made in reference to the Status Quo. Once gone the West would view an attack, if not directed against itself, then at minimum as an attack of its values – something that would make Western engagement considerable more forceful. If it is right that prospects of losses and gains influence players' strategic deliberation, then this is the sort of mistakes that we risk making when adhering to standard game theoretic assumptions. Instead of guiding they lead deterrence analysis astray (Krepinevich 2019).

It has been argued that in order to adjust for how agents actually make choices prospect theoretical findings not only improve analysis but also redefine theories of deterrence since it brings agents' reference points as well as their biases into the equation. Consequently, it is argued that we need to adjust our theorizing of deterrence to correct for prospect theoretical findings (Berejikian

CONTACT Karl Sörenson  karl.sorenson@fhs.se  KTH - Royal Institute of Technology, FHS - Swedish National Defence University, Drottning Kristinas väg 37, PO-BOX 2780, Stockholm 115935, Sweden

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

1997, 2002; Haas 2001; McDermott 1998; Levy 1996, 1997, 2002). Several scholars have developed this line of theorizing and advanced the arguments for modifying theories that originally relied on standard game theoretic assumptions so that they can account for references and biases players might have (Hafner-Burton, Haggard, Lake and Victor 2017; Morrisette 2010).²

Parallel to the discussion about prospect theory's potential role in deterrence theory, some of the more central concepts in PT such as loss-aversion have been reconsidered by behavioral economists (Harrison and Don 2017; Gal and Rucker 2018). The critique does not question PT as a whole, but shows that some of the reported effects may not always exist. Hence, we are in a position to state something more comprehensive about the benefits and shortcomings of deterrence that relies on PT assumptions. The claim that a deterrence theory would benefit from relying on PT results seems straightforward: changing agents' possible ideal behavior to agents' actual behavior would make for an empirically grounded theory of deterrence. The questions explored in this article are: Do prospect theoretical results change how we should view deterrence? And, do we need to construct a new theory of deterrence as a consequence of prospect theoretical results?

While this article treats deterrence theory within international security and whether or to what extent prospect theoretical inferences should be a part of such a theory, there are similar discussions in related fields of research. Hafner-Burton et al. (2017) suggest that it is no less than a 'behavioral revolution in international relations'. Jason Morrisette (2010) theorizes about how prospect theory could be the foundation for 'a theory of cognitive realism'. While these two articles depart from deterrence and how this sort of dynamics is affected by prospect theoretical findings, the main claim would potentially affect other research that has a similar relation between behavior models and theories of a behavior. Thus, there is a principal argument with regard to inferences drawn from prospect theory and its relation to theories in the social sciences.

The article unfolds as follows: first, the basic claims of a theory of deterrence, as understood when relying on standard assumptions, are mapped out. The article then mathematically shows how the strategies in a strategic game change when prospect theoretical assumptions are applied. From these results the article proceeds by appraising the changes and fit them into previous research on deterrence and prospect theory and connect these to the more general discussion about the model-theory relationship, thereby summarizing the contribution of prospect theory to deterrence research and deterrence theory.

Deterrence Representation, Theory and Evidence

As stated in the introduction, several scholars have turned to prospect theory in attempts to show that strategies of deterrence should be understood in a different way. The studies of Hafner-Burton et al. (2017) and of Morrisette (2010) theorize about how prospect theory potentially could make for a more cognitive realistic understanding of international conflicts. These two studies draw upon an influential article entitled 'A Cognitive Theory of Deterrence' by Jeffrey Berejikian (2002). The studies are interesting as they open up theorizing about deterrence to psychological factors that may be a part of the decision-making.

Berejikian and his followers argue that a model functions as the foundation for a theory.³ This view is not unique. Stephen Walt (1999), Stephen Quackenbush (2011a), Fred Lawson (2013) and Frank Zagare (1999, 2013) all explicitly argue from a similar point: which deterrence model represents deterrence best? Zagare and Quackenbush are proponents of the Perfect Deterrence models, several games that share the same game-form of Bayesian games with different player types. This theory and the games associated with it is however criticized by Walt and Lawson, who believe that the classic understanding of deterrence reflected in the game Chicken, or the Mutual Deterrence game is for better or worse inescapable if one wants to understand the dynamics of deterrence.

The Mutual Deterrence game is also the game Berejikian principally relies on when proposing 'a cognitive theory of deterrence'. From the Mutual deterrence game many conjectures regarding deterrence have been made, Tomas Schelling (1960, 1966) and Herman Kahn (1960) are perhaps the

most notable, but the game has worked as starting point for a plethora of security political analyses that focus on deterrence in one form or another. Zagare and Kilgour, but also Powell (1990), formulate their games in response to the perceived shortcomings with the Mutual Deterrence game (Sörenson 2017).

The Mutual Deterrence game is a normal form game. A normal form game, G , consists of a number of *players*, in this case two, $P = (\text{player1}, \text{player2})$ the strategy space S_i for each player i , and *payoff function*, $u_i \in U$, which gives the von Neumann-Morgenstern utility $u_i(s)$ for every strategy $s_i \in S$.

In the Mutual Deterrence game each player prefers to attain an Advantage (outcome DC for player 1; CD for player 2) over the Status Quo outcome (CC), which in turn is preferred to a Disadvantage (CD for player 1; DC for player 2) and Conflict (DD) is the least preferred outcome. There are two pure strategy Nash equilibria: one where one player plays cooperate, strategy C , and the other player plays defect, strategy D ; and its compliment where the other player cooperates and the first defects. This is because if one player believes that the other player defects its best reply is to cooperate. If

		Player 2	
		C	D
Player 1	C	0, 0	-1, 1
	D	1, -1	-2, -2

Preference order:

Player 1 := $DC > CC > CD > DD$

Player 2 := $CD > CC > DC > DD$

Figure 1. Chicken/Mutual Deterrence Game.

a player believes that the other player will cooperate, its best reply is to the defect.

Chicken also has a mixed strategy equilibrium where p represents the probability that player 1 cooperates and q the probability that player 2 cooperates:

$$(CC)(q) + (DC)(1 - q) = (CD)(q) + (DD)(1 - q)$$

$$(CC)(p) + (CD)(1 - p) = (DC)(p) + (DD)(1 - p)$$

Given the payoffs displayed in Figure 1., $p = q = 1/2$, hence, the players are indifferent between the strategies cooperate and defect.⁴ The dilemma of Mutual Deterrence captures a central aspect with mutual deterrence, namely that attempting an Advantage, strategy D , risks Conflict, while choosing strategy C , to cooperate, risks that the opponent takes advantage (see Brams and Kilgour 1988 for a more thorough analysis).

The Mutual Deterrence game is thought to represent an important part of the Cold War dynamics. Deterrence is achieved when one of the players backs down because of the other player's choice to defect. Alternatively, when players mix their strategies, they are likely to cooperate to a relatively high degree. It is this latter type of interaction that might be said to represent mutual deterrence. While the Mutual Deterrence game undoubtedly is a central game Berejikian also discusses Entry Deterrence as an equally important (Berejikian 2002). The game, first proposed by Reinhard Selten

(1978), extends the Mutual Deterrence to a sequential game, where one player is challenging and the other is defending. In the game there are two equilibria, one where the Challenger remains in the Status quo and one where when Challenger attacks and Defender yields. When studied with subgame perfect equilibrium, there is no deterrence. This is because if Challenger attacks, the rational choice for the Defender is to yield, since yielding is preferred to conflict by Defender. Selten's game shows that credible threats are necessary for deterrence (Selten 1978; Radner and Rosenthal 1982, but also Zagare 2018).

Deterrence scholars have had various solutions to the problem of credibility. To Schelling deterrence is a question of brinkmanship, where the player that is willing to venture closest to the brink deters the opponent (Schelling 1960, 199).⁵ Similarly, Kahn discusses the idea of pre-commitment as a method for a player to deter an opponent (Kahn 1960, 146–147). Pre-commitment is a solution that recognizes the problem with credible threats, but this solution is not a game theoretical solution (see Quackenbush 2011b, 7).

As mentioned, a later but central development is the so-called *Perfect Deterrence*. The authors Zagare and Kilgour use a game-form from which they formulate Perfect Mutual Deterrence, a mutual deterrence game – that replaces the Mutual deterrence game, Perfect Asymmetric Deterrence, which replaces Selten's sequential deterrence game, and Extended Deterrence, where a third party protects a smaller player from an antagonist. The games assume two types of players, those with credible threats, and those without credible treats. Credible threats are equated with threats that are believable and a threat is believable if the player has incentives to carry it out. A credible player therefore has a different preference order where it prefers Conflict to Disadvantage (e.g. for player 1 this would mean $DC > CC > DD > CD$ (Zagare and Kilgour 1993, 4). Hence, Perfect Deterrence covers several central deterrence dynamics.

In a study by Stephen Quackenbush (2010); Quackenbush (2011b), the Perfect Asymmetric Deterrence game is statistically tested. The results make Quackenbush conclude that the Perfect Deterrence model accurately predicts when deterrence fails and when it succeeds (Quackenbush 2010, 2011b).⁶ Hence, if we are interested in understanding asymmetric deterrence, then there is a strong case to be made for the Perfect Asymmetric Deterrence game.

The Perfect Asymmetric Deterrence game has also been explored from a PT perspective (Carlson and Dacey 2006). In their study the authors substitute the standard expected utility function with a PT function. The authors investigate the claim that deterrence failure is due to a declining value of the Status Quo. They find that this certainly can be the case but not always – not even when a Challenger is operating under a losses frame and views its loss more negative than it actually is (Carlson and Dacey 2006, 196). In sum, the Perfect Asymmetric Deterrence game has been corroborated statistically with the standard assumptions as well as tested theoretically from a PT perspective.

The Perfect Asymmetric Deterrence game presents an important case since the theoretical construct of the model can be shown to hold for the studied empirical examples, and as shown by Carlson and Dacey, the dynamics can be susceptible to player biases. Taken together the Perfect Asymmetric Deterrence game seems to be a model that theoretically holds and corresponds to known cases of conventional deterrence.

The situation is however more complicated with regard to mutual deterrence. This is because these games are meant to capture nuclear deterrence similar to that of the Cold War. Therefore these games cannot be empirically validated in the same sense since the outcomes of the Cold War and later nuclear deterrence interaction all have a similar outcome. The games are to some extent conjectures and assumptions regarding the type of interaction mutual deterrence seem to entail. This is why PT could be important. It relocates the empirical base from the actual game to the decision-making. The Mutual Deterrence game is in this context central as it is a game that most deterrence theorists in one way or another revert to, like Berejikian (2002), incorporate, as Zagare and Kilgour (2000) or at minimum depart from, like Powell (1990) (see also Sörenson (2017 who discusses the game's foundational value to deterrence).

So what type of changes can we expect if we introduce PT assumptions into deterrence modeling? The claim that prospect theoretical results require a reconstruction of a deterrence theory implies that something fundamental in the structure changes. Changes in a model can be of different types. Zagare and Kilgour discuss how the Mutual deterrence game's strategic situation changes into a different game, the Prisoner's dilemma, when players' have credible threats (Zagare and Kilgour 2000, 76). The aforementioned Carlson and Dacey (2006) study focuses on changes in the equilibria. When such changes occur, one needs to appraise how many of the equilibria that are affected and what this implies if one is trying to construct a theory based on this model. Butler (2007) who analyzes Fearon's bargaining game, akin to some of the deterrence games, reaches similar conclusions as the Carlson–Dacey study. Such distinctions will help to classify how large a type of change actually is. In general, a quantitative change in a player's payoffs constitutes one type of change. A quantitative change in a player's payoffs that affects the preference order can have considerably larger implications. While it is central to note what type of change an alternation may cause, it is equally important to contextualize what such a change means for the game and what implications this may have for a theory. Thus, the type of change as well as its meaning must be addressed when settling the question: What type of changes does PT induce for a theory of deterrence?

Prospects of Deterrence

Jervis (1976) suggests that deterrence, just as many other forms of strategic communication, is context sensitive to the agents involved. On a similar note, Snyder (1976) points out that there are psychological factors that may account for deterrence dynamics. The prospect theoretical project attempts to account for such deviation systematically by studying how people tend to make their decisions when faced with a certain type of choices. Prospect theory assigns a value function v and a probability weighted function w . Where $v(x)$ reflects the subjective value, which is the objective outcome y relative to a reference r , where $x = y - r$.

$$v(x) = \begin{cases} x^\beta & x \geq 0 \\ -\lambda(-x)^\beta & x < 0 \end{cases} \quad (1)$$

The value function is summarized as '(i) defined on deviations to the reference point; (ii) generally concave for gains and commonly convex for losses; (iii) steeper for losses than for gains' (Kahneman and Tversky 1979, 278).

The weight function states that agents tend to overweight small probabilities and underweight moderate to high probabilities. The probabilities $w(p) + w(p) + \dots w(p)$ do not always sum to 1. The weighting function is expressed in (Kahneman and Tversky 1992)⁷:

$$w(p) := \frac{p^\gamma}{(p^\gamma + 1 - p)^\frac{1}{\gamma}} \quad (2)$$

Where $w(p)$ = weighted probability. Prelec axiomatizes a sub-proportional function that satisfies the properties of the probability distribution that simplifies the analysis (Prelec 1998, 499):

$$w(p) := e^{-(\ln p)^\alpha} \quad (3)$$

Where $0 < \alpha < 1$. The function has a fixed point and an inflection point at $p = 1/e = .37$ (see Figure 3). Prelec suggests that $\alpha = .65$, which further simplifies the analysis. For β and λ , the empirically reported values of .88 and 2.25, respectively, are used in the analysis.

Taking the value function and weighted probability function together the value of x with probability p and the value of y with probability q is expressed as:

$$v_i(x; p; y; q) = w(p)v(x) + w(q)v(y) \quad (4)$$

In this sense a behavior of risk-aversion when faced with a gain and a behavior of risk-seeking when faced with a loss is captured by PT. Because PT report behavior of an agent making a monetary gambling decision, it is not entirely straightforward how this translates to game theory. Metzger and Rieger note that the probabilities that are transformed by the probability-weighting function complicates the analysis of the mixed strategies in Nash equilibrium considerably since ‘... [the] objective probability with which one player chooses a particular strategy will be transformed to a (different) subjective probability by another player who will choose his own strategy according to this subjective probability, rather than the objective probability.’ (Metzger and Rieger 2009, 4). This aspect needs to be taken into account in an analysis of deterrence behavior.

There is no theory of how nation states view prospects of winning and losing in deterrence situations (Levy 1997, 191). However, there are several case studies (McDermott 1998; Haas 2001; Morrisette 2010) and more general types of conjectures; Berejikian (1997, 2002) Levy (1997; 2002), Carlson and Dacey (2006), and Butler (2007) discuss how players may regard their prospects in a strategic environment. As mentioned, in their analysis of player reference points, Carlson and Dacey tests the idea of how critical a Challenger’s value of the Status Quo is (Carlson and Dacey 2006). Butler, drawing on Levy, discusses how other reference points such as aspirations and social comparison between actors may influence an agent’s reference point (Levy 2003, 270; Butler 2007, 232). Similar for both discussions are that they are mindful of the games they are analyzing and choosing the reference points accordingly.

Central is that the value of a player’s options are made relative to a reference point, r . Where $r_i \in [0, 1]$ for $i \in \{1, 2\}$. This means that if x is an expected value to player i than this is appraised relatively to the reference point, $x - r_i$ for player i . By varying the reference point for the players we can study various scenarios (Shalev 2002; Butler 2007).⁸

Deterrence Analysis under Prospect Theoretical Assumptions

With pure strategies, PT does not change the dynamics of the Mutual deterrence game, since the monotonic increase/decrease does not affect the preference order for the player. Hence, the focus of the analysis is the mixed strategies. The question is which reference points that are most relevant or interesting to study. One reference point that is of general interest to investigate is when the players orientate themselves from the *Status Quo*. This is the point of departure for many deterrence discussion, both Levy and Butler discuss it and Carlson and Dacey uses it as a point of departure. However, the Status Quo also is considered more widely in PT.

Another reference point of interest is to see what happens when a player has the *disadvantage* outcome as reference, whereas the other has the conflict outcome as reference point. The third case to consider is when both players have the *advantage* as reference point. The purpose of this study is not only to see what happens to a particular game that is analyzed with the aid of PT but to investigate if there are changes in the equilibrium between the different scenarios that suggest that standard analysis is inadequate and that a prospect theoretical foundation might make more sense than the standard assumptions. Therefore more substantial deviations are of interest. We depart from the payoffs displayed in Figure 1 and relate the results to the game with standard assumptions where $p = q = 1/2$.

Case 1: Status Quo Bias?

Status Quo is a central concept in deterrence theory. It is often viewed as a default position from which a potential challenger may or may not make a first move. In extended form games such as Entry deterrence (Selten 1978), Nuclear deterrence (Powell 1990), Perfect asymmetric/symmetric deterrence (Zagare and Kilgour 2000) the Status Quo is one of the choices a challenger chooses between at the outset of the game. As mentioned, Carlson and Dacey (2006) make the value of the Status Quo as the point of departure for their prospect theoretical analysis. In PT, Status Quo is analysed as a type of bias where ‘the Status Quo choice acts as a psychological anchor’, i.e. an agent

is inclined to remain in the Status Quo since it at times does not involve making an active decision at all (Samuelson and Zeckhauser 1988, 41). Further, Samuelson and Zeckhauser point out that the bias also seems to apply to organizations (Samuelson and Zeckhauser 1988, 41 see also Daniel, Knetsch, and Thaler Richard 1991; Levy 1996). The importance of the Status Quo outcome is therefore something that is worth investigating – at minimum to see if the PT-analysis of a strategic form game confirms the intuitions about its effects on player behavior. If both players have their reference points set to the Status Quo, i.e. $r_1 = CC$ and $r_2 = CC$, then both players should be more likely to play cooperate. As before p represents the probability that player 1 plays cooperate and q represents the probability that player 2 plays cooperate.

$$(CC - r_1)w(p) + \lambda(CD - r_1)^\beta w(1 - p) = (DC)^\beta(w)p + \lambda(DD - r_1)^\beta w(1 - p)$$

and

$$(CC - r_2)w(q) + \lambda(CD - r_2)^\beta w(1 - q) = (CD)^\beta(w)q + \lambda(DD - r_2)^\beta w(1 - q)$$

With Status Quo as the reference point for both players the scenario corroborates the idea of player behavior as more risk averse. Both players are relatively more reluctant to defect, i.e. to play strategy D , and are more likely to play cooperate, strategy C , to a higher degree, see Figure 3. This is a deviation from the game under standard assumptions. It also confirms the effect reported by PT of having Status Quo as a reference point.

The Status Quo bias is not merely a theoretical possibility, but a distinct problem that is related to deterrence. Chan (1979), identifies the Status Quo bias within intelligence agencies as a severe obstacle for successful strategic warning. This observation is corroborated with the analysis above as Status Quo bias will anchor a player to playing cooperate with a higher degree. However, Jervis (2010) continues this discussion by also considering the difficulties within agencies to identify biases an opponent may operate under. In the analysis above, the players can identify the bias in each other; however, this may not always be possible, as pointed out by Jervis.

Case 2: Mixed Frames

It is likely that players have divergent points of reference. This means that they weigh the same strategic situation in different ways. I consider a scenario where player 1's reference point is the disadvantage outcome (CD) $r_1 = CD$ and that player 2's reference point is the conflict outcome (DD) $r_2 = DD$. By doing so we can explore both what these reference points imply in terms of strategic outlook for the players, but also how the dynamic changes when the reference points are different ($r_1 \neq r_2$).

$$(CC - r_1)^\beta w(p) + (CD - r_1)^\beta w(1 - p) = (DC - r_1)^\beta(w)p + (DD - r_1)^\beta w(1 - p)$$

and

$$(CC - r_2)^\beta w(q) + (DC - r_2)^\beta w(1 - q) = (CD)^\beta(w)q + (DD - r_2)^\beta w(1 - q)$$

With these reference points player 1 is marginally more likely to play strategy C than strategy D . Player 2 is also more likely to play strategy C than D . The reference point that makes it most likely to play strategy C , is when the opponent has the disadvantage outcome (CD for player 1, and DC for player 2). The result that is most divergent from deterrence under standard assumptions (i.e. without PT assumptions) is when a player reference point is disadvantage (see Figure 3).

Case 3: Mutual Possible Destruction

The most central scenario that in many ways is what motivates research on deterrence is the possibility of the conflict outcome – the possibility of a full nuclear exchange between the rivaling

actors. From Russell (1959), Kahn (1960) and Schelling (1960) to Powell (1990) and Zagare and Kilgour (2000) the central problem is to capture and appraise under which conditions this scenario might occur. So what does Prospect Theory have to say about this? By putting the reference point to the advantage position for each player ($r_1 = DC$ and $r_2 = CD$), the value of the game for both players is lowered.

$$\lambda(CC - r_1)^\beta w(p) + \lambda(CD - r_1)^\beta w(1 - p) = (DC - r_1)^\beta (w)p + \lambda(DD - r_1)^\beta w(1 - p)$$

and

$$\lambda(CC - r_2)^\beta w(q) + \lambda(DC - r_2)^\beta w(1 - q) = (CD - r_2)^\beta (w)q + \lambda(DD - r_2)^\beta w(1 - q)$$

In the eventuality that both players have their best respective outcomes as reference point, they are both more likely to play defect and – as a consequence, this is the scenario where it is most likely that both players get their most preferred outcome – or least preferred outcome (given that the opponent plays the same strategy in the latter case). This is also the most extreme scenario with the highest probability that both players play defect (Figure 3). USSR's decision to move nuclear missiles to Cuba was precipitated by the US' decision to move nuclear missiles into Italy and Turkey, spawning the Cuban missile crisis. It could be argued, given the analysis above that it was the focus on attaining an advantage from both actors that ended up putting them on a trajectory that could have ended in Armageddon.

Figure 2, shows the normalized expected value of player i in mixed equilibrium as a function of r . Payoffs are normalized by subtracting the utility from the Status Quo outcome given the reference point r . This is done to mitigate the mechanical effect of the reference point on the expected value from any outcome as a higher reference point always leads to a lower utility for the players. The figure shows how the expected value changes depending on the reference point if it is normalized for $r_i = CC$, the Status Quo outcome. The expected value of the game increases up to $r =$ the Disadvantage outcome (CD for player 1, DC for player 2) along with the increase of r . The expected value then decreases when r increases.

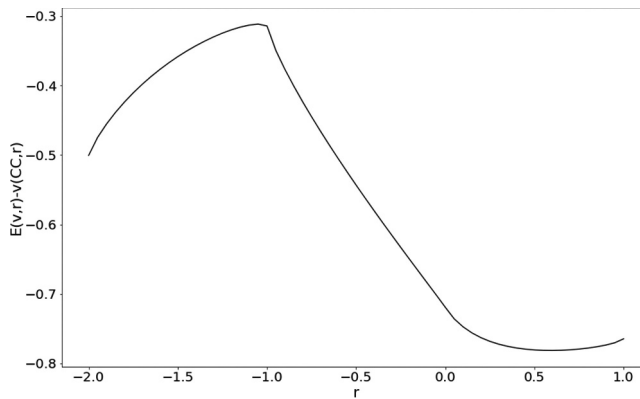


Figure 2. Normalized expected value of player i in mixed equilibrium as a function of r .

Figure 3, shows how the reference point changes the dynamics of the game. The probability (q) of cooperation increases from $r =$ Conflict (DD) to $r =$ Disadvantage (DC). Under these conditions, a player is more likely to cooperate than to defect. The probability, q , then decreases from $r =$ Disadvantage (DC) to $r =$ Status Quo (CC), but still remain above $1/2$, i.e. $q > .5$. This is the region

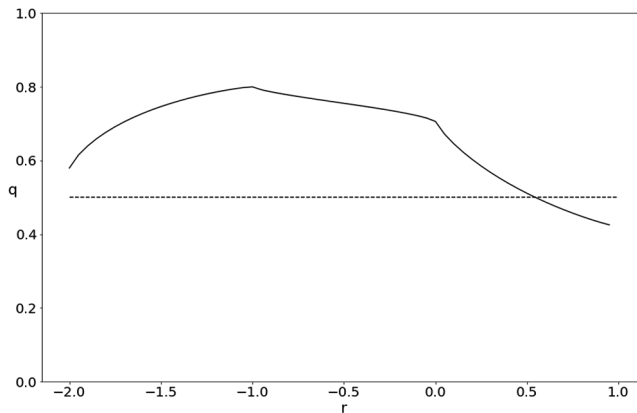


Figure 3. Equilibrium probability q of cooperation as a function of r .

where we find the first case, the so-called Status Quo Bias scenario, described above. Between the reference point $r = \text{Status Quo}$ and the Advantage outcome (CD), q continues to decrease. Central is that between $r = \text{Status Quo}$ and $r = \text{Advantage}$, the probability to defect increases. When a player's reference point is between Status Quo and Advantage a player is more likely to play defect than to play cooperate, i.e. $q \leq .5$. This reference point is the reference points for a player under the 'Mutual Possible Destruction' scenario described above.

So how do these results compare to earlier research on conflict modeling with PT assumptions? The differences found when analysing the Mutual Deterrence game are in many ways similar to those found in the Carlson and Dacey (2006) and the Butler study (2007), as well as the Mazicioglu and W (2018) study. Prospect theory changes the dynamics of the games, sometimes quite clearly, but not in a fundamental way. Given how the reference point functions, the order of preferences is never affected, but the weight of the various preferences are. This impacts the mixed strategies, and in the Mutual Deterrence game the changes are clearly there, but the changes does not alter the preference order. The differences between the Mutual Deterrence game, the Carlson and Dacey (2006) and the Butler study (2007) is largely due to that the two latter are Bayesian games with player types with different preference orders, in the first, and in the second case, Fearon's bargaining game with a divisible good. This makes these games more susceptible to changes when analysed with PT.⁹

Hence, to the extent that there are changes induced by PT analysis, they corroborate the robustness of the dynamics of the Mutual deterrence game in general, but also show that given a specific context PT assumption may be important. However, our questions do not only pertain to whether PT leads to differences, but if such changes require a change to the fundamentals of a deterrence theory.

The House That Prospect Theory Built

One of the first scholars to discuss PT in an international relations context was Levy, who pointed out that the results from PT were promising, but that more work needed to be done in order to solidify the results (Levy 1997, 191). Harrison and Don (2017) analyses the main issues that have been brought up against prospect theory. Central to their analysis is that the experiments, which are the foundation for PT are difficult to reproduce (Harrison and Don 2017).¹⁰ While the research community does not seem to be able to replicate the results exactly, there are several experiments that arrive at similar conclusions, but with variations (Camerer 2003; Weber and Johnson 2009). To what degree we should have confidence in that PT reports real behavior is therefore not entirely clear, only

that given the proper context the measures seem to be a reasonable estimate. In light of this type of evidence, or rather lack thereof, the type of research that simply applies the PT to deterrence, should precede with caution since some of the reported effects may not exist or be exaggerated. Suggesting the reconstruction of an entire theory based on such research is overstating what can actually be claimed.

It has been pointed out that one of the reasons that game theory (in general) has not gained more in popularity is because it is difficult to interpret what the mixed strategies signify, i.e. when players randomize between their strategies (Radner and Rosenthal 1982, 401). In deterrence research game theory is prevalent, and there are clearly strong epistemic reasons for this.¹¹ However, if the mixed strategy analysis feels contra-intuitive with standard assumption, the complexity increases when analysed through PT. This may be acceptable if we felt convinced that it represented a good way to capture actual behavior. PT is based on decisions in binary monetary choice-situations, hence translating it into utility and strategic interaction may lose some of the representational quality.

Further, other types of experiments have found that PT is outperformed by other types of boundedly rational models such as the so-called *Rank-Dependent Utility* (RDU) (Harrison and Don 2017). Yet another alternative approach would be to focus on how people behave when they play a game for the first time. Crawford (2003) uses the so-called level-*k* model to capture how players represent or misrepresent their intentions.¹² The point is not that level-*k* or RDU would be even better foundations than standard game theory (or PT) to base a deterrence theory on, merely that it is not clear why PT should take precedence over other behavior model with empirical support.

A model that aims to replace the standard assumptions need to show how we should appraise and apply such new knowledge. For modeling a particular situation using PT assumptions can clearly be justified. Carlson and Dacey's investigation of declining Status Quo and deterrence failure, is clearly a case where applying PT is both motivated and relevant. However, if the focus is not a particular situation, or like in the case of the Carlson–Dacey answering a specific question, but to reconstruct a theory of deterrence, then a better type of motivation how the statistical underpinnings of PT and why PT should be viewed is crucial. For a cognitive theory of deterrence, stating the case for the choice of PT and refuting competing models should be important.

A Little Less False Theory of Deterrence?

Till Grüne-Yanoff points out that so-called bounded rational models often only mean a slight improvement of the models that they are meant to replace and 'at best gives us theories that are a little less false' (Grüne-Yanoff 2007). So, if this is true, is a change away from the standard assumptions worth it? Is it a little less false? When summing up the research that combines prospect theoretical insights and applications to deterrence there is a discernible separation between how the research community relates to models and theory. The construct of the model usually accounts for how something specific functions, but from the model we also learn something much more general, which goes beyond the specific target of the model (Sugden 2000, 5). From this point of view, there is a whole range of situations that a model potentially can account for.¹³ This is also why several theorists of science have concluded that models provide an interesting foundation for theory. Not only because it is the models within a theory that capture the phenomenon a theory aims to generalize about, it is also because much of the scientific enterprise's focus is on the construction and testing of models (see for instance Giere 2004; Weisberg 2013). By emphasizing models this school of thought does make sense when viewing how deterrence theory has developed.

The discussion between Zagare, Walt, Lawson and Quackenbush show that the Mutual Deterrence game is central for deterrence theorizing, but that it plays a contested role. In contemporary deterrence theories it is the starting point, but confined to a stipulated space or excluded from the actual theorizing. In Perfect Deterrence theory, the Mutual deterrence game is a particular instance of Perfect Mutual deterrence when both players lack credible threats – and in Powell's Nuclear Deterrence it is the reference and starting point from which insights are deduced, but then

abandoned to construct something new (Powell 1990, 35). This is because for contemporary scholars such as Zagare, Kilgour, Powell, Slantchev and Quackenbush the relation between model and theory is essential. Their work is defined by critique of earlier models that motivate the development of their own models from which theory or theoretical discussions are explored. It is central how the new models lay the foundation for the dynamics that they later theorize about.

Followers such as Quackenbush (2010); Quackenbush (2011b) also pay close attention to the representation and that it is coherent with any deductions that may follow. The fact that the Perfect Asymmetric deterrence model is statistically corroborated strengthens the case for this type of representation for conventional asymmetric deterrence. In addition, the PT analysis by Carlson and Dacey of the Perfect Asymmetric Deterrence model shows that such assumptions make a difference, but as in the case with the Mutual Deterrence game, the changes do not affect the general dynamics. That is, it does not put the general construct of the Perfect Asymmetric Deterrence game into question. Rather, it is a compliment when trying to answer a particular question or understand a specific instance (see Gigerenzer 2002 on models as heuristic or toolbox). Hence, the Perfect Asymmetric Deterrence game and its associated theory are not affected by the prospect theoretical results, but instances in modeling may be.

This is similar to what is found when analyzing the Mutual Deterrence game. The changes that occur when one applies PT do not motivate the type of changes that put the general dynamics of the game into question, e.g. the players' preference orders; it simply alters the probability distribution within the range. Like the Perfect Asymmetric deterrence game, a switch to PT assumptions can be motivated when trying to answer a particular question or mapping specific situation, but it does not alter theoretical deductions already reached.

Overarching the actual application of PT to deterrence is a more principal argument. If one views the model-theory relation as fundamental, as the more contemporary school of thought tend to do, then the consequences of changing the utility function are profound as rational play will no longer be center for how we analyze player interaction. The idea to ground a deterrence theory, like Perfect Deterrence theory or Nuclear Deterrence, on how people actually behave would seriously limit such a theory of deterrence. The point of game theoretic analysis is to think of a situation and strip away all the noise that historical records may imply to only analyze what rational players would do. Such knowledge is essential when building a theory of deterrence, because no matter how a context and type of players may change, we need to know the potential situation of players optimizing in order to build a full-bodied theory. In this sense rational behavior provide an upper bound for the theory. If PT was made foundational, thereby shifting the base of deterrence to empirical records, it would narrow the situation to actual behavior at best and put the deterrence theorizing on the wrong path at worst.

It is possible to take a less stern view of the theory–model relationship and rather focus on the plurality of models. We can then conceive of the plurality of various deterrence models as different perspectives on deterrence. Co-existing models is an important discussion. Mikaela Massimi (2018) suggests that models relate either through *plurality*, i.e. several of the models are in play in a given scientific context, by *partiality* – that each of them provides a partial account for the phenomenon, or they are *complementary*, i.e. their heuristic function is that they, according to specific rules that vary between scientific contexts, jointly map a phenomenon (Massimi 2018, 15). With such an approach PT assumption may well exist alongside rational representations and other forms of bounded rational models. Massimi is not alone in this approach, Walter Veit (2020) also argues for such a pluralistic approach to economic modeling. PT would then simply become an additional perspective on deterrence behavior. This 'perspectival' approach to models is in many ways an attractive position, as it allows for a range of theorizing that does not exclude certain aspects. However, the problem with this more open-ended view is that it makes the contours of a deterrence theory blurry. When a whole range of models that may run contrary to one and other together make up a theoretical landscape new contributions will be difficult to evaluate since it is unclear what theory they are contributing to.

This last aspect is probably why many scholars that set out proposing a new deterrence theory with a PT foundation end up discussing the possibilities that open up to remodel the *phenomenon of deterrence* rather than what it changes in a *deterrence theory* it supposedly is a contribution to. When focus is shifted from theory to phenomenon, PT as replacement for standard assumptions is lost and it becomes a compliment for capturing a phenomenon. This is a reasonable view to take, but then the theoretical contribution to a deterrence theory remains unclear and one would still have to accept that for developing theory, PT can never become a replacement for rational play.

So, if PT does not change anything fundamental with the dynamics of deterrence, what is its contribution? One conceivable answer is the more precise theorizing about nation state's reference points, biases and frames. Carlson and Dacey, Butler, Mazicioglu and Merrick, as well as this study, assume alternative reference points based on reasonable conjectures. Levy and Berejikian have shown what the principal mechanisms of biases and frames are central – it is what they lead to that is less clear. A first step would therefore be to develop quantifiable criteria how a state forms its references. Such a step would not only help the modeling, but also be a contribution to a deterrence theory since it would make a state's biased reference points more than an open discussion.

Conclusion

Could prospect theory better explain why actors at times make decisions that seem sub-optimal? Why agents take greater risks than is deemed rational? And if so, at what level of analysis should such insight be put to use? The question asked in the beginning of this article was whether prospect theoretical results change how we should view deterrence, and if we need to reconstruct a theory of deterrence based on PT? We are now in a position to reply. The answer to the first question is, potentially yes. The mathematical analysis of the Mutual Deterrence game under PT assumptions presented in this article shows that the changes are at times quite clear. The Cuban missile crisis is potentially a case where such aspects may shed additional light on how the crisis came to a head. Thus, in certain cases prospect theoretical aspects can be of use when appraising a strategic situation. This reply comes with an important caveat: as things currently stand empirical evidence for the effects suggested by PT are not entirely straightforward. Hence, even when convinced of an agent's actual bias, the extent to which such a bias such be applied is unclear.

The reply to the second question, if we should reconstruct a theory of deterrence, must be in the negative. The marginal changes, the indefinite status of experiments that PT rests upon, the unclear prioritization of PT over other types of bounded rational models, are all premises in the argument against reformulating a new theory of deterrence based on PT. While PT can potentially be helpful to answer a specific question or model a particular situation, it cannot supplant the need to analyze a phenomenon idealized, like in classic game theory, since this is central for the development of deterrence theory. In fact, it would be detrimental for a deterrence theory to exclude what optimal play would look like.

The benefits of PT seem on the one hand to be on a case-to-case basis. If it is known that an actor operates with a different reference, PT could facilitate a more nuanced analysis. On the other hand, the benefit of PT is on a structural level since it takes strategic deliberation and cognitive processes into account, which game theory does not do. While PT cannot replace standard game theoretical assumption, it can help us to reason about what potentially constrains an agent's reasoning - in theory and in practise.

Notes

1. Indeed, this is the type of analysis the French intelligence agency seems to have made, appraising the probability of a Russian attack as low, i.e. unlikely (France24 2022-03-31).
2. These claims must however be separated from studies that explore how deterrence dynamics change when a prospect theoretical utility function is applied instead of the standard von Neumann-Morgenstern utility function. Carlson and Dacey (2006) are in this context central and a study this article in part relies upon. In

a similar vein Butler (2007); Mazicioglu and W (2018) explore game theoretic differences and similarities between standard and prospect theoretic assumptions.

3. One of the first scholars to consider prospect theory is Levy (1996, 1997, 2002). Levy's approach to prospect theory has a quite general aim as he discusses various aspects that are relevant to international relations, but not always directly pertaining to deterrence.
4. The Cold War scenario is often depicted by lowering the Conflict outcome, DD , which changes the dynamics so that the players are less willing to play defect
5. Schelling's solution has spawned several debates, see (Field 2014) who question its relevance and Zagare's reply (2018) to Field and Sörenson's (2022) reply to Zagare and Field.
6. Quackenbush uses dyads to define cases of deterrence and lets the outcomes of the game be the dependent variable. Quackenbush controls for balance of forces between the rivaling states, foreign policy, geographical proximity, democracy and peace years. In this, the more statistically driven some of the more central contributions include (Huth and Russett 1984; Lebow and Gross Stein 1990a; Huth and Russett 1988; Lebow and Gross Stein 1990b; Liberman 1994).
7. PT was further developed to Cumulative Prospect Theory, which is primarily concerned with the probability weighing function and adds a clear differentiation between probabilities associated with gains and probabilities referring to losses, in Cumulative Prospect Theory the probabilities do sum to 1 (Kahneman and Tversky 1992).
8. The reference point can be used to define what commonly is referred to as 'the frames'. An agent with a reference point that makes it view its prospect more negative is said to operate under a losses frame, whereas an agent with a relatively more positive reference is said to be acting under a gains frame (see for instance Levy 2003). However, for a strategic game such as Chicken, it is not entirely clear how such frames can be defined, given that the reference point is associated with one outcome, which defines the values for the remaining outcomes, according to (1). Hence, for this analysis it is sufficient and more stringent to only use the reference point.
9. How the results compare to Berejikian's analysis of the Mutual Deterrence game is difficult to assess since he relies on an informal approach and disregards mixed strategies.
10. Harrison and Ross' point is not that Kahneman's and Tversky's conclusions are wrong, but rather that 'anybody casually using these estimates as statistically representative must not care about rigour in empirical work' (Harrison and Don 2017, 154). See also Gal and Rucker (2018) who point out that that new experimental findings regarding loss aversion suggests that the original results have been over-interpreted.
11. If one wants to analyse how two or more actors' choices impact each other in terms of outcomes and strategy choices, game theory is the natural choice, however there are other approaches. For instance Jeffrey (1983) considers a type of deterrence game from a decision theoretic perspective and Freedman (2004) points out that criminologists also study deterrence, but with a partly different focus than political scientists. An early example of such a study is Landes (1978) who makes a economic study of aircraft high-jackings with the aid of PT.
12. Haruvy and Stahl (2005) tested selection criteria that people rely on when they play a game for the first time. They corroborated previous studies (see for instance Stahl 1993) that there was a difference in the cognitive depth between test-subjects, which they systematized into levels of cognitive sophistication.
13. 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).

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Karl Sörenson  <http://orcid.org/0000-0003-1939-9392>

News articles

France24. 2022-03-31. 'French military spy chief quits after failure to predict Russian invasion' (<https://www.france24.com/en/france/20220331-french-military-spy-chief-quits-after-failure-to-predict-russian-invasion>)

References

- Berejikian, J. D. 1997. "The Gains Debate: Framing State Choice." *American Political Science Review* 91 (4): 789–805. doi:10.2307/2952164.
- Berejikian, J. D. 2002. "A Cognitive Theory of Deterrence." *Journal of Peace Research* 39 (2): 165–183. doi:10.1177/0022343302039002002.
- Brams, S., and M. Kilgour. 1988. *Game Theory and National Security*. New York, USA: Blackwell Publishing.
- Butler, C. K. 2007. "Prospect Theory and Coercive Bargaining." *Journal of Conflict Resolution* 51 (2): 227–250. doi:10.1177/0022002706297703.
- Camerer, C. F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. New Jersey USA: Princeton University Press.
- Carlson, L., and R. Dacey. 2006. "Sequential Analysis of Deterrence Games with a Declining Status Quo." *Conflict Management and Peace Science* 23 (2): 181–198. doi:10.1080/07388940600666022.
- Chan, S. 1979. "The Intelligence of Stupidity: Understanding Failures in Strategic Warning." *The American Political Science Review* 73 (1): 171–180. doi:10.2307/1954739.
- Crawford, V. P. 2003. "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions." *American Economic Review* 93 (1): 133–149. doi:10.1257/000282803321455197.
- Dalsjö, R., M. Jonsson, and J. Norberg. 2022. "A Brutal Examination: Russian Military Capability in Light of the Ukraine War." *Survival* 64 (3): 7–28. doi:10.1080/00396338.2022.2078044.
- Daniel, K., J. L. Knetsch, and H. Thaler Richard. 1991. "The Endowment Effect, Loss Aversion and the Status Quo Bias." *Journal of Economic Perspectives* 5 (1): 193–206. doi:10.1257/jep.5.1.193.
- Field, A. J. 2014. "Schelling, von Neumann, and the Event that Didn't Occur." *Games* 5 (1): 53–89. doi:10.3390/g5010053.
- Gal, D., and D. Rucker. 2018. "The Loss of Loss Aversion: Will It Loom Larger than Its Gain?" *Journal of Consumer Psychology* 28 (3): 497–516. doi:10.1002/jcpy.1047.
- Giere, R. N. 2004. "How Models are Used to Represent Reality." *Philosophy of Science* 71 (5): 742–752. doi:10.1086/425063.
- Gigerenzer, G. 2002. "The Adaptive Toolbox." In *Bounded Rationality: The Adaptive Tool Box*, edited by R. Selten and G. Gigerenzer, 37–50, Cambridge, Massachusetts, USA.: MIT Press.
- Grüne-Yanoff, T. 2007. "Bounded Rationality." *Philosophy Compass* 2 (3): 534–563.
- Haas, M. L. 2001. "Prospect Theory and the Cuban Missile Crisis." *International Studies Quarterly* 45 (2): 241–270. doi:10.1111/0020-8833.00190.
- Hafner-Burton, E. M., H. Stephen, D. A. Lake, and D. G. Victor. 2017. "The Behavioural Revolution and International Relations." *International Organisation* 71 (1): 1–31. doi:10.1017/S0020818316000400.
- Harrison, G. W., and R. Don. 2017. "The Empirical Adequacy of Cumulative Prospect Theory and Its Implications for Normative Assessment." *Journal of Economic Methodology* 24 (2): 150–165. doi:10.1080/1350178X.2017.1309753.
- Haruvy, E., and D. O. Stahl. 2005. "Equilibrium Selection and Bounded Rationality in Symmetric Normal-form Games." *Journal of Economic Behaviour and Organization* 62 (1): 98–119. doi:10.1016/j.jebo.2005.05.002.
- Huth, P., and B. Russett. 1984. "What Makes Deterrence Work? Cases from 1900 to 1980." *World Politics* 36 (4): 496–526. doi:10.2307/2010184.
- Huth, P., and B. Russett. 1988. "Deterrence Failure and Crisis Escalation." *International Studies Quarterly* 32 (1): 29–45. doi:10.2307/2600411.
- Jeffrey, R. 1983. *The Logic of Decision*. Chicago, USA: University of Chicago Press.
- Jervis, R. 1976. *Perception and Misperception in International Politics*. New Jersey, USA: Princeton University Press.
- Jervis, R. 2010. *Why Intelligence Fails - Lessons From the Iranian Revolution and the Iraq War*. Ithaca, New York USA: Cornell University Press.
- Kahn, H. 1960. *On Thermonuclear War*. New Brunswick, New Jersey, USA: Princeton University Press, reprinted by Transaction Publishers 2007.
- Kahneman, D., and A. Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (, 2): 263–291. doi:10.2307/1914185.
- Kahneman, D., and A. Tversky. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5 (4): 297–323. doi:10.1007/BF00122574.
- Krepinevich, A. J. F. 2019. "The Eroding Balance of Terror." *Foreign Affairs*, no. 98: 62–74.
- Landes, W. M. 1978. "An Economic Study of U. S. Aircraft Hijacking, 1961–1976." *The Journal of Law and Economics* 21 (1): 1–19. doi:10.1086/466909.
- Lawson, F. 2013. "Back to the Future in the Study of Deterrence." *St Anthony's International Review* 9 (1): 144–156.
- Lebow, R., and J. Gross Stein. 1990a. *When Does Deterrence Succeed and How Do We Know?* Ottawa, Canada: Canadian Institute for International Peace and Security.
- Lebow, R., and J. Gross Stein. 1990b. "Deterrence: The Elusive Dependent Variable." *World Politics* 42 (3): 336–369. doi:10.2307/2010415.
- Levy, J. S. 1996. "Loss Aversion, Framing and Bargaining: The Implication of Prospect Theory for International Conflict." *International Political Science Review* 17 (2): 179–195. doi:10.1177/019251296017002004.

- Levy, J. S. 1997. "Prospect Theory, Rational Choice and International Relations." *International Studies Quarterly* 41 (1): 87–112. doi:[10.1111/0020-8833.00034](https://doi.org/10.1111/0020-8833.00034).
- Levy, J. S. 2002. "Application of Prospect Theory to Political Science." *Synthese* 135 (2): 215–241. doi:[10.1023/A:1023413007698](https://doi.org/10.1023/A:1023413007698).
- Lieberman, E. 1994. "The Rational Deterrence Theory Debate: Is the Dependent Variable Elusive?" *Security Studies* 3 (3): 384–427. doi:[10.1080/09636419409347556](https://doi.org/10.1080/09636419409347556).
- Massimi, M. 2018. "Perspectival Modelling." *Philosophy of Science* 85 (3): 335–359. doi:[10.1086/697745](https://doi.org/10.1086/697745).
- Mazicioglu, D., and M. J. R. W. 2018. "Behavioural Modelling of Adversaries with Multiple Objectives in Counterterrorism." *Risk Analysis* 38 (5): 962–977. doi:[10.1111/risa.12898](https://doi.org/10.1111/risa.12898).
- McDermott, R. 1998. *Risk-Taking in International Politics: Prospect Theory in American Foreign Politics*. Michigan, USA: University of Michigan Press. Ann Arbor.
- Metzger, L. P., and M. O. Rieger. 2009. "Equilibria in Games with Prospect Theory Preferences." *National Centre of Competence in Research Financial Valuation and for Risk Management*, no. 598.
- Morrisette, J. J. 2010. "Rationality and Risk-Taking in Russia's First Chechen War: Toward a Theory of Cognitive Realism." *European Political Science Review* 2 (2): 187–210. doi:[10.1017/S1755773910000081](https://doi.org/10.1017/S1755773910000081).
- Powell, R. 1990. *Nuclear Deterrence – The Search for Credibility*. Cambridge, UK: Cambridge University Press.
- Prelec, D. 1998. "The Probability Weighing Function." *Econometrica* 66 (3): 497–527.
- Quackenbush, S. L. 2010. "General Deterrence and International Conflict: Testing Perfect Deterrence Theory." *International Interactions* 36 (1): 60–85. doi:[10.1080/03050620903554069](https://doi.org/10.1080/03050620903554069).
- Quackenbush, S. L. 2011a. "Deterrence Theory: Where Do We Stand?" *Review of International Studies* 37 (2): 741–762. doi:[10.1017/S0260210510000896](https://doi.org/10.1017/S0260210510000896).
- Quackenbush, S. L. 2011b. *Understanding General Deterrence: Theory and Application*. New York, USA: Palgrave MacMillan.
- Radner, R., and R. Rosenthal. 1982. "Private Information and Pure Strategy Equilibria." *Mathematics of Operations Research* 7 (3): 401–409. doi:[10.1287/moor.7.3.401](https://doi.org/10.1287/moor.7.3.401).
- Rubinstein, A. 2006. "Dilemmas of an Economics Theorist." *Econometrica* 74 (4): 865–883. doi:[10.1111/j.1468-0262.2006.00689.x](https://doi.org/10.1111/j.1468-0262.2006.00689.x).
- Russell, B. 1959. *Common Sense and Nuclear Warfare*. Simon and Schuster.
- Samuelson, W., and R. Zeckhauser. 1988. "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty* 1 (1): 7–59. doi:[10.1007/BF00055564](https://doi.org/10.1007/BF00055564).
- Schelling, T. C. 1960. *The Strategy of Conflict*. Cambridge, Massachusetts, USA: Harvard University Press.
- Schelling, T. C. 1966. *Arms and Influence*. New Haven, USA: Yale University Press.
- Selten, R. 1978. "The Chain Store Paradox." *Theory and Decision* 9 (2): 127–159. doi:[10.1007/BF00131770](https://doi.org/10.1007/BF00131770).
- Shalev, J. 2000. "Loss Aversion Equilibrium." *International Journal of Game Theory* 29: 269–287.
- Snyder, J. 1976. "Rationality on the Brink – The Role of Cognitive Processes in Failures of Deterrence", RAND Paper Series.
- Sörenson, K. 2017. "Comparable Deterrence: Target, Criteria and Purpose." *Defence Studies* 17 (2): 198–213. doi:[10.1080/14702436.2017.1321468](https://doi.org/10.1080/14702436.2017.1321468).
- Sörenson, K. 2022. "A Misfit Model: Irrational Deterrence and Bounded Rationality", *Theory and Decision*, (online – forthcoming).
- Stahl, D. O. 1993. "Evolution of Smartn Players." *Games and Economic Behavior* 5 (4): 604–617. doi:[10.1006/game.1993.1033](https://doi.org/10.1006/game.1993.1033).
- Sugden, R. 2000. "Credible Worlds: The Status of Theoretical Models in Economics." *Journal of Economic Methodology* 77 (1): 1–31. doi:[10.1080/135017800362220](https://doi.org/10.1080/135017800362220).
- Veit, W. 2020. "Model Pluralism." *Philosophy of the Social Sciences* 50 (2): 91–114. doi:[10.1177/0048393119894897](https://doi.org/10.1177/0048393119894897).
- Walt, S. 1999. "Rigor or Rigor Mortis? Rational Choice and Security Studies." *International Security* 23 (4): 5–48. doi:[10.1162/isec.23.4.5](https://doi.org/10.1162/isec.23.4.5).
- Weber, E. U., and E. J. Johnson. 2009. "Decisions under Uncertainty: Psychological, Economic, and Neuroeconomic Explanations of Risk Preference." In *Neuroeconomics*, edited by W. Glimcher Paul, C. F. Camerer, and Fehr. Russell A: Ernst and Poldrack.
- Weisberg, M. 2013. *Simulation and Similarity – Using Models to Understand the World*. Oxford, UK: Oxford University Press, 127–144.
- Zagare, F. 1999. "All Rigor, No Mortis." *International Security* 24 (2): 107–114. doi:[10.1162/016228899560185](https://doi.org/10.1162/016228899560185).
- Zagare, F. 2013. "Deterrence Theory Then and Now: There Is No Going Back." *St Anthony's International Review* 9 (1): 157–167.
- Zagare, F. 2018. "Explaining the Long Peace: Why von Neumann (And Schelling) Got It Wrong." *International Studies Review* 20 (3): 422–437. doi:[10.1093/isr/vix057](https://doi.org/10.1093/isr/vix057).
- Zagare, F., and M. Kilgour. 1993. "Asymmetric Deterrence." *International Studies Quarterly* 37 (1): 1–27. doi:[10.2307/2600829](https://doi.org/10.2307/2600829).
- Zagare, F., and M. Kilgour. 2000. *Perfect Deterrence*. Cambridge, UK: Cambridge University Press.