

Summer 2014

Does the Prisoner's Dilemma Refute the Coase Theorem?, 47 J. Marshall L. Rev. 1289 (2014)

Enrique Guerra-Pujol

Orlando Martínez-García

Follow this and additional works at: <http://repository.jmls.edu/lawreview>

 Part of the [Law and Economics Commons](#), and the [Law and Philosophy Commons](#)

Recommended Citation

Enrique Guerra-Pujol & Orlando Martínez-García, Does the Prisoner's Dilemma Refute the Coase Theorem?, 47 J. Marshall L. Rev. 1289 (2014)

<http://repository.jmls.edu/lawreview/vol47/iss4/8>

This Article is brought to you for free and open access by The John Marshall Institutional Repository. It has been accepted for inclusion in The John Marshall Law Review by an authorized administrator of The John Marshall Institutional Repository.

DOES THE PRISONER'S DILEMMA REFUTE THE COASE THEOREM?

ENRIQUE GUERRA-PUJOL¹
ORLANDO I. MARTÍNEZ-GARCÍA²

I. Introduction	1290
II. Standard Versions of the Prisoner's Dilemma	1291
A. Numerical Form	1291
B. Algebraic or Logical Form	1295
III. Coasean Version of the Dilemma (with Strategic and Non- Strategic Bargaining)	1297
A. A Tale of Two Parables: Parable of the Rancher and the Farmer and Parable of the Prisoners	1298
B. The Three Conditions of the Coase Theorem	1300
1. Reciprocal Nature of the Prisoner's Dilemma	1300
2. Well-Defined Property Rights	1301
3. Zero Transaction Costs, Strategic Behavior, and Non-Strategic Bargaining	1302
4. Strategic Bargaining, Threats and Promises in the Prisoner's Dilemma	1303
5. Non-Strategic Coasean Bargaining	1304
IV. The Role of Uncertainty, Exponential Discounting, and Elasticity in the Coasean Version of the Prisoner's Dilemma	1306
A. Uncertainty	1306
B. Exponential Discounting	1307
C. Price Elasticity of Demand	1309
Example #1	1312
Example #2	1312
Example #3	1313
D. Lessons and Discussion	1313
V. A Brief Digression Regarding the Role of Third Parties in the Prisoner's Dilemma	1314
VI. Some Closing Thoughts on the Complexity of the Prisoner's Dilemma	1316
VII. Conclusion	1317

¹ Lecturer, University of Central Florida, Dixon School of Accounting. College of Business Administration. J.D., Yale Law School. B.A., UCSB. Enrique Guerra-Pujol presented a previous draft of this paper on the morning of December 7, 2013 at the 2013 Northeast People of Color Legal Scholarship Conference held at the University of Puerto School of Law and wishes to thank Taja-Nia Y. Henderson, Orlando Martinez-Garcia, Adys Ann Guerra, Sydja Robinson, Hamed G. Santaella, Carlitos del Valle, and Judge Jenny Rivera for attending my talk and for their helpful comments and suggestions.

² Visiting Professor, Interamerican University of Puerto Rico Law School. Adjunct, University of Puerto Rico, Arecibo Campus, Department of Social Science. LL.M., University of Pennsylvania. J.D., Interamerican University of Puerto Rico Law School. B.A., The American University.

		II	
		confess	not confess
I	confess	(-1, -1)	(1, -2)
	not confess	(-2, 1)	(0, 0)

3

I. INTRODUCTION

Building upon the main theme of this year's LatCrit Conference, *Resistance Rising: Theorizing and Building Cross-Sector Movements*,⁴ this paper (i.e., our contribution to this larger critical conversation) challenges one of the dominant paradigms in economics and law: the Coase Theorem.⁵ Specifically, we present a thought-experiment, what we shall call the "pure Coasean version" of the famous Prisoner's Dilemma game.⁶ In brief, what if the prisoners in this game-theory parable were allowed to communicate and bargain with each other instead of being held in separate cells, as in the standard version of the dilemma? Would our prisoners strike a mutually-beneficial and collectively-optimal Coasean bargain, as the Coase Theorem predicts?⁷ Or, as predicted in the standard one-shot version of the Prisoner's Dilemma in which bargaining is not allowed,⁸ would they still end

³ A.W. Tucker, *A Two-Person Dilemma: The Prisoner's Dilemma* (1950), as reprinted in Philip D. Straffin, Jr., *The Mathematics of Tucker: A Sampler*, 14 TWO-YEAR C. MATHEMATICS J. 228 (1983).

⁴ Latina & Latino Critical Legal Theory, Inc., *2013 Biennial LatCrit Conference Program Schedule (and Related Events)*, (2013), available at http://latcrit.org/media/medialibrary/2013/10/LatCrit2013_Conference_Program_FinalR.pdf.

⁵ The Coase Theorem is named after the late Ronald Coase. Ronald H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1, 1-44 (1960). George Stigler, however, was the economist who first presented the idea now known as the Coase Theorem. GEORGE J. STIGLER, *THE THEORY OF PRICE* 113 (MacMillan, 3d ed. 1966). George Stigler stated Coase's idea as a "theorem" and coined the term "Coase Theorem." *Id.*

⁶ See generally WILLIAM POUNDSTONE, *PRISONER'S DILEMMA* (Anchor Books 1993) (providing an overview and history of the origins of the dilemma); see also F. E. Guerra-Pujol, *The Parable of the Prisoners*, 5-9 (June 21, 2013) [hereinafter Guerra-Pujol, *The Parable of the Prisoners*] (unpublished manuscript) (on file with author), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2281593, (explaining the prisoner's parable).

⁷ *Id.*

⁸ See *infra* Part I.B.

up defecting?

Before proceeding, it is worth noting that few scholars have explored the possible relation between the Coase Theorem and the Prisoner's Dilemma. One important exception is Wayne Eastman, a professor at Rutgers Business School, who established a formal identity between the Coase Theorem and the Prisoner's Dilemma.⁹ Instead of following Eastman's approach (i.e., relating the Coase Theorem to the Prisoner's Dilemma),¹⁰ we do the opposite. We relate the Prisoner's Dilemma to the Coase Theorem by constructing a pure Coasean version of the dilemma.

The remainder of this paper is organized as follows: Part II provides some background by presenting the standard formulations of the Prisoner's Dilemma in numerical as well as algebraic terms. Next, Part III presents our thought-experiment: in order to test the true value of the Coase Theorem, we consider a "pure Coasean version" of the Prisoner's Dilemma in which property rights are well-defined and transactions costs are zero (i.e., the prisoners are allowed to openly communicate and bargain directly with each other). Part IV explores the effects of (i) uncertainty, (ii) exponential discounting, and (iii) elasticity on the behavior of the prisoners in the Coasean version of the dilemma. Part V considers the role of the prosecutor (and third parties, generally) in the Prisoner's Dilemma and the overall complexity of the dilemma. Lastly, Part VI identifies conditions under which the Prisoner's Dilemma refutes the Coase Theorem, while Part VII concludes.

II. STANDARD VERSIONS OF THE PRISONER'S DILEMMA

By way of background, we begin this paper by presenting the standard or "canonical" formulation of the Prisoner's Dilemma – by far the most famous story or parable in all of game theory – both in numerical and algebraic form.¹¹ Readers who are already familiar with the details of the Prisoner's Dilemma may skip this part and proceed to Part III.

A. Numerical Form

The original formulation of the dilemma is attributed to

⁹ See Wayne Eastman, *How Coasean Bargaining Entails a Prisoners' Dilemma*, 72 NOTRE DAME L. REV. 89, 95–98 (1964) (establishing a formal identity between the Coase Theorem and the Prisoner's Dilemma).

¹⁰ *Id.* at 90 n.7.

¹¹ See sources cited *supra* note 6 and accompanying text.

Professor Albert Tucker, a mathematician at Princeton University, who presented the parable of the prisoners during a guest lecture at Stanford University in May 1950.¹² Specifically, Professor Tucker posed the following hypothetical scenario in a one-page mimeograph titled *A Two-Person Dilemma* that he prepared for his guest lecture:

- Two men, charged with a joint violation of law, are held separately by the police. Each is told that
- (1) if one confesses and the other does not, the former will be given a reward of one unit and the latter will be fined two units,
 - (2) if both confess, each will be fined one unit.
- At the same time each has reason to believe that
- (3) if neither confesses both will go clear.¹³

In addition, Professor Tucker included the following “payoff table” in his mimeo to illustrate his parable:

		II	
		confess	not confess
I	confess	(-1, -1)	(1, -2)
	not confess	(-2, 1)	(0, 0)

14

Although Professor Tucker does not use the terms “Prisoner’s Dilemma” or “Prisoners’ Dilemma” in his original mimeo, he does refer to the prisoners’ predicament as “a two-person dilemma” in the title of the mimeo.¹⁵ More importantly, in Tucker’s original telling of his tale, we see all the elements associated with the standard version of the Prisoner’s Dilemma:¹⁶

-Two Suspects: *I* and *II* are held in separate rooms and thus unable to communicate or bargain with each other;

-Two Choices: confess or not confess;

-Interdependent Payoffs: the payoffs associated with each choice depend upon the choices made by both suspects;

-Payoff Table: a visual presentation of the parable, or stated formally, a reduction of the dilemma to “normal form.”

The first published account of the Prisoner’s Dilemma, however, does not appear until several years later, when R. Duncan Luce and Howard Raiffa’s published their classic game-

¹² SYLVIA NASAR, *A BEAUTIFUL MIND: A BIOGRAPHY OF JOHN FORBES NASH, JR.* 118 (1998); POUNDSTONE, *supra* note 6 at, 117–18.

¹³ See Tucker, *supra* note 3 (presenting the parable of the prisoner).

¹⁴ Tucker, *supra* note 3.

¹⁵ *Id.*

¹⁶ See *id.* (presenting all of the essential elements of the standard versions of the Prisoner’s Dilemma model).

theory treatise, *Games and Decisions: Introduction and Critical Survey*:¹⁷

The following interpretation [of a two-person, non-zero-sum game], known as the prisoner's dilemma, is popular: Two suspects are taken into custody and separated. The district attorney is certain that they are guilty of a specific crime, but does not have adequate evidence to convict them at trial. He points out to each prisoner that each has two alternatives: to confess to the crime the police are sure that they have done, or not to confess. If they both do not confess, then the district attorney states he will book them on some very minor trumped-up charge such as petty larceny and illegal possession of a weapon, and they will both receive minor punishment; if they both confess they will both be prosecuted, but he will recommend less than the most severe sentence; but if one confesses and the other does not, then the confessor will receive lenient treatment for turning state's evidence whereas the latter will get "the book" slapped at him.¹⁸

In addition, Luce and Raiffa express the payoffs of their prisoners' parable in numerical form (i.e., in terms of years in prison) in a payoff table, stating that "the strategic problem" in this particular parable "might reduce to" the following set of payoffs:

		Prisoner 2	
		Not Confess	Confess
Prisoner 1:	Not Confess	1 year each	10 years for 1 and 3 months for 2
	Confess	3 months for 1 and 10 years for 2	8 years each

¹⁹

It is worth noting that Luce and Raiffa specifically included Professor Tucker's strategic game and their own corresponding payoff matrix in the chapter devoted to "Two-Person Non-Zero-

¹⁷ R. DUNCAN LUCE & HOWARD RAIFFA, *GAMES AND DECISIONS: INTRODUCTION AND CRITICAL SURVEY* (Dover Publications 2012) (1957).

¹⁸ *Id.* at 95.

¹⁹ *Id.*

Sum Non-Cooperative Games.”²⁰ In so doing, Luce and Raiffa present the Prisoner’s Dilemma parable in order to illustrate a particular model of strategic behavior – what game theorists refer to as a “two-person, non-zero-sum, non-cooperative game.”²¹ This standard version of the Prisoner’s Dilemma thus encompasses all the elements essential to such *two-person, non-zero-sum, non-cooperative games*:

-First and foremost, the Prisoner’s Dilemma is a simple *two-person* model or game; there are only two prisoners. This is an important simplifying assumption, since there could just as well be three, four, or *n* number of suspects. By reducing the number of players in this parable to just two suspects, it simplifies the underlying strategic situation.

-Second, the Prisoner’s Dilemma is a *non-zero-sum* game insofar as both suspects can receive light sentences if both are able to remain silent instead of snitching. In a zero-sum game, by contrast, the gain of one player always comes at the expense of the other player. Moreover, in a non-zero-sum game, such as the Prisoner’s Dilemma, a “win-win” outcome is possible, but only if both players agree to cooperate with each other.

-Next, the prisoners in this story are playing a *non-cooperative* game. The prisoners are *incommunicado* insofar as they are held in separate cells to prevent them from bargaining with each other. Strictly speaking, a non-cooperative game rules out the possibility of mutually beneficial Coasean bargaining among the players. (For our part, we shall later modify this aspect of the Prisoner’s Dilemma when we present our pure Coasean version of the dilemma in part two.)

-Last, but not least, the Prisoner’s Dilemma is a *one-shot* game: the prisoners have only one opportunity to play the game. Although this requirement is not stated explicitly in Luce and Raiffa’s interpretation of the parable, subsequent research has shown that cooperation is theoretically possible when the game is played many times (iteration) and when the occurrence of the last round is uncertain.²²

²⁰ *Id.* at 94–97.

²¹ See generally, John F. Nash, *Non-Cooperative Games*, 54 ANNALS OF MATHEMATICS 286 (1951). Luce and Raiffa, however, were one of the first to express this particular type of game in the form of the Prisoner’s Dilemma. LUCE & RAIFFA, *supra* note 17. As a further aside, John von Neuman and Oskar Morgenstern also presented non-zero-sum games in their foundational game theory treatise, but the focus of their work is on cooperative games (i.e. games in which bargaining among the players is allowed), not on non-cooperative games, such as the Prisoner’s Dilemma. JOHN VON NEUMANN & OSKAR MORGENSTERN, *THEORY OF GAMES AND ECONOMIC BEHAVIOR* 504–86 (Princeton Univ. Press, 3d ed. 1953).

²² See generally ROBERT AXELROD, *THE EVOLUTION OF COOPERATION* 11 (Basic Books rev. ed. 2006) (1984) (exploring various resolutions to the Prisoner’s Dilemma when one player in this two-person game plays the game

In addition, Luce and Raiffa attribute this standard interpretation of the Prisoner's Dilemma to A.W. Tucker and also note that this example "has received considerable attention by game theorists."²³ That this particular parable was already "popular" by the mid-1950s – and sufficiently well-known among mathematicians to be included in Luce and Raiffa's treatise on game theory – is itself telling. But why did this parable become so popular so quickly? One possible reason is the realism of the Prisoner's Dilemma.

Simply put, Luce and Raiffa's version of this parable seems to capture the legal system "in action," or, more specifically, how the criminal justice system actually operates when the prosecution does not have sufficient evidence to go to trial, much less convict an individual.²⁴ Briefly, when he is stymied by a lack of evidence, the prosecutor must adjust his strategy, for without the cooperation of at least one of the prisoners, he will only be able to secure a conviction on some "minor trumped-up charge" (to borrow Luce and Raiffa's phrasing). And thus it should come as no surprise that the tactics of offering "lenient treatment" for cooperation (i.e., getting a suspect to "flip" or turn State's evidence) and of filing "trumped-up charges" (what criminal defense attorneys refer to as "overcharging") are common strategies used by prosecutors.²⁵

B. Algebraic or Logical Form

Thus far, we have presented the standard Prisoner's Dilemma in *numerical* form, but the payoffs in this model can also be presented in *algebraic* or logical form. Consider, for example, the following payoff table, which presents the Prisoner's Dilemma in both numerical and algebraic form:

with many iterations – "with an indefinite number of interactions, cooperation can emerge"); see also Robert Axelrod & William D. Hamilton, *The Evolution of Cooperation*, 241 SCIENCE 1390, 1391–92 (1981) (presenting a model based on the concept of an evolutionarily stable strategy in the context of the Prisoner's Dilemma game).

²³ LUCE & RAIFFA, *supra* note 17, at 94.

²⁴ *Id.* at 94–97.

²⁵ *Id.* at 95.

		Player B	
		C Cooperation	D Defection
Player A	C Cooperation	R=3 Reward for mutual cooperation	S=0 Sucker's payoff
	D Defection	T=5 Temptation to defect	P=1 Punishment for mutual defection

26

Here, we shall focus on the descriptive labels *C* and *D* and the algebraic labels *R*, *S*, *T*, and *P*. First, the players' choices or "strategy sets" of the players in this matrix now appear in more general terms: "cooperation" ("*C*") and "defection" ("*D*") correspond to "confess" and "not confess," respectively, in the traditional version of the Prisoner's Dilemma.²⁷ Likewise, the payoffs are now designated in general terms. For example, "Reward" ("*R*") represents the payoff for mutual cooperation, "Punishment" ("*P*") represents the payoff for mutual defection, and "Temptation payoff" ("*T*") and "Sucker's payoff" ("*S*") represent the remaining two payoffs.²⁸

Stated formally (i.e., in general algebraic terms as opposed to specific numerical values), a game is a Prisoner's Dilemma when the values of the payoffs are ranked in ordinal fashion: $T > R > P > S$.²⁹ Moreover, regardless of whether the Prisoner's Dilemma is presented in numerical or algebraic form, the outcome and logic of this game remain the same: defection is always a dominant strategy, or "Nash equilibrium,"³⁰ in the one-shot version of the game.

If the other player cooperates, there is a choice between cooperation which yields *R* (the reward for mutual cooperation) or defection which yields *T* (the temptation to defect). By assumption, $T > R$, so that it pays to defect if the other cooperates. On the other hand, if the other player defects, there is a choice between cooperation which yields *S* (the sucker's payoff) or defection which

²⁶ Axelrod & Hamilton, *supra* note 22, at 1392.

²⁷ *Id.* at 1391–92.

²⁸ *Id.* For our part, we shall follow this convention and thus continue to refer to the payoffs in the Prisoner's Dilemma using these standard labels in these algebraic or general terms.

²⁹ See Axelrod & Hamilton, *supra* note 22, at 1392, Figure 1 (identifying that the "payoff to Player A is shown with illustrative numerical values[, and] [t]he game is defined by $T > R > P > S$ and $R > (S + T) / 2$ ").

³⁰ See, e.g., SYLVIA NASAR, A BEAUTIFUL MIND: A BIOGRAPHY OF JOHN FORBES NASH, JR. 15, 20 (1998).

yields P (the punishment for mutual defection). By assumption $P > S$, so it pays to defect if the other player defects. Thus, no matter what the other player does, it pays to defect. But, if both defect, both get P rather than the larger value of R that both could have gotten had both cooperated. Hence, the dilemma.

With two individuals destined to never meet again, the only strategy that can be called a solution to the game is to defect always despite the seeming paradoxical outcome that both do worse than they could have had they cooperated.³¹

Therefore, whether the parable is presented in numerical or algebraic form, the central lesson of the standard one-shot version of the Prisoner's Dilemma is that defection or "snitching" is always the Nash equilibrium of the game. Moreover, from an individual perspective, both prisoners are always better off by defecting, regardless of the other prisoner's actions. For example, if the other prisoner ("Player B") snitches, Player A might as well snitch to avoid S , the sucker's payoff. In fact, even if the other prisoner keeps quiet, Player A is still better off snitching insofar as T , the temptation payoff, is always greater than R .

But what if the prisoners are not held in separate rooms (i.e., they are not *incommunicado*)? What if the prisoners could actually bargain with each other and had the ability to make enforceable promises and credible threats? Would they still defect? We shall consider these questions next by presenting a "Coasean version" of the Prisoner's Dilemma, one in which the prisoners are allowed to communicate and bargain with each other.

III. COASEAN VERSION OF THE DILEMMA (WITH STRATEGIC AND NON-STRATEGIC BARGAINING)

The previous section discussed the standard version of the Prisoner's Dilemma, in which both players are separated with no means to communicate with each other. This section, however, removes the element of separation and presents a theoretical test of the Coase Theorem through a novel thought-experiment – a pure Coasean version of the Prisoner's Dilemma, one in which bargaining and communication are allowed between the prisoners. First, we compare the standard version of the Prisoner's Dilemma with the Coase Theorem, identifying the main conditions of the theorem: (i) the existence of a "reciprocal" conflict between two parties; (ii) well-defined property rights; and (iii) zero transaction costs. Next, we explain how our pure Coasean version of the

³¹ Axelrod & Hamilton, *supra* note 22, at 1391.

Prisoner's Dilemma satisfies these conditions, specifically considering the application of strategic as well as non-strategic bargaining in the Coasean or zero transaction cost version of the dilemma.

A. *A Tale of Two Parables: Parable of the Rancher and the Farmer and Parable of the Prisoners*

Broadly speaking, two of the most important ideas in economics and law are the Coase Theorem and the Prisoner's Dilemma, and each has generated a vast technical literature – a scholarly sea of Borgesian proportions.³² And yet, each of these profoundly influential contributions is based on a simple parable: Ronald Coase's "Parable of the Rancher and the Farmer"³³ and the game-theoretic "Parable of the Prisoners."³⁴ In summary, Coase's parable concerns two pastoral neighbors, a cattle rancher and a crop farmer,³⁵ while Tucker's tale involves two criminal suspects apprehended by the police.³⁶ Although these memorable parables evoke wildly different and divergent worlds (i.e., a bucolic world of neighboring farms and ranches on the one hand versus a *film noir* world of cops and robbers on the other), from an economic perspective, these simple stories share an essential facet in common. In brief, both parables depict rational actors whose interests collide. In the one case, the conflict arises out of cattle trespass (i.e., the rancher's cattle trampled the farmer's crops); in the other, each prisoner must decide whether to betray or remain loyal to the other.³⁷

Nevertheless, although both parables portray rational parties with opposing or conflicting interests, these stories diverge in one

³² At last count (July 25, 2014), for example, an electronic search for the term "Coase theorem" generates 25,200 results. GOOGLE SCHOLAR, http://scholar.google.com/scholar?q=%22coase+theorem%22&btnG=&hl=en&as_sdt=0%2C14 (last visited July 25, 2014). A search for the term "prisoner's dilemma," however, produces more than four times as many results (107,000). *Id.* at http://scholar.google.com/scholar?q=%22prisoner%27s+dilemma%22&btnG=&hl=en&as_sdt=0%2C14 (last visited July 24, 2014).

³³ See Coase, *supra* note 5, at 2–15 (presenting the "Parable of the Rancher and the Farmer"). See also F. E. Guerra-Pujol, *Modelling the Coase Theorem*, 5 EUR. J. LEGAL STUD. 139, 141–42 (2012) [hereinafter Guerra-Pujol, *Modelling*] (combining Coase's intuitive insights with the formal methods of game theory); Robert Ellickson, *Of Coase and Cattle: Dispute Resolution Among Neighbors in Shasta County*, 38 STAN. L. REV. 623, 624–25 (1986) (reporting the results of an attempt to explore the realism of the "Parable of the Rancher and the Farmer").

³⁴ LUCE & RAIFFA, *supra* note 17; Guerra-Pujol, *The Parable of the Prisoners*, *supra* note 6; Tucker, *supra* note 3.

³⁵ Coase, *supra* note 5, at 2–15.

³⁶ Tucker, *supra* note 3.

³⁷ *Id.*; Coase, *supra* note 5, at 2–15.

important respect: the ability, or lack thereof, of the parties to settle their differences through bilateral negotiations or Coasean bargaining. That is, the most salient distinction between the hypothetical worlds of the Coase Theorem and the Prisoner's Dilemma is the ability to bargain. In the former case, the rancher and the farmer are fully able to bargain with each other and negotiate a mutually beneficial enforceable agreement.³⁸ In the latter story, however, the prisoners have no such option; they are held in separate cells and unable to talk, much less bargain with one another.³⁹ This ability, or inability, of the parties to deal with each other is of critical importance, at least in the traditional telling of each tale. In the one case, a Coasean bargain between the rancher and the farmer produces an optimal result or Panglossian outcome (i.e., an efficient allocation of resources devoted to the production of crops and meat).⁴⁰ In the other case, the parties' inability to bargain with each other inevitably leads to mutual betrayal and a worse outcome (longer prison sentences) for both prisoners.⁴¹

Suffice it to say, however, few scholars have explored the relation between these two important models. One exception is Wayne Eastman, a professor at Rutgers Business School, who identified the conditions under which Coasean bargaining constitutes a Prisoner's Dilemma.⁴² Specifically, he models Coase's rancher-farmer parable as a Prisoner's Dilemma and establishes a formal identity between the Coase Theorem and the Prisoner's Dilemma.⁴³ Our approach in this paper, however, is different than Eastman's. Instead of relating the Coase Theorem to the Prisoner's Dilemma, as Eastman does, we do the opposite.⁴⁴ We relate the Prisoner's Dilemma to the Coase Theorem by constructing a Coasean version of the dilemma.⁴⁵ Specifically, we pose the following questions: what if the prisoners were, in fact, allowed to communicate and bargain with each other in Coasean fashion? That is, what if our hapless prisoners were able to negotiate a mutually beneficial and legally enforceable agreement? Would

³⁸Coase, *supra* note 5, at 2–15.

³⁹ LUCE & RAIFFA, *supra* note 17; Tucker, *supra* note 3.

⁴⁰ Coase, *supra* note 5, at 2–15.

⁴¹ Tucker, *supra* note 3.

⁴² Eastman, *supra* note 9.

⁴³ *Id.*

⁴⁴ See *id.* at 90 n.7 (noting very deliberately that his proposition “relates the [Coase] Theorem to the [Prisoner's] Dilemma, rather than vice versa” and his reasons for electing to do so).

⁴⁵ See Guerra-Pujol, *Modelling*, *supra* note 35 (providing a different game-theoretic formulation of the Coase Theorem).

they still defect or would they somehow decide to cooperate as postulated by the Coase Theorem?

B. *The Three Conditions of the Coase Theorem*

Before attempting to answer the above questions, we shall first identify and review the main conditions of the Coase Theorem. Professor Coase introduced the counterintuitive idea now known as the “Coase Theorem” with a memorable parable about cattle trespass.⁴⁶ The rancher-farmer parable, however, is really a story about joint interactions involving bargaining and property rights. That is, Coase posed a well-defined reciprocal problem using the example of cattle trespass and then imagined what would happen if the affected parties (i.e., the rancher and the farmer) could solve this problem through voluntary bargaining.⁴⁷ (Ultimately, this is the same question that we pose about the prisoners in the Prisoner’s Dilemma.) Coase observed that when (i) the costs of transacting are zero (a standard assumption in economics) and (ii) property rights are well-defined, “Coasean bargaining” (i.e., voluntary negotiations) among the affected parties will produce an efficient economic outcome.⁴⁸ Although this economic “theorem” has been stated in many different ways over the years,⁴⁹ the necessary elements of the Coase theorem remain constant: (i) the existence of a reciprocal conflict, (ii) well-defined property rights, and (iii) zero transactions costs (i.e., no impediments to bargaining).⁵⁰ Accordingly, we shall now show how our pure Coasean version of the dilemma satisfies all three standard conditions of the Coase Theorem.

1. Reciprocal Nature of the Prisoner’s Dilemma

First and foremost, we wish to point out the “reciprocal nature”⁵¹ of the prisoners’ plight in all versions of the Prisoner’s Dilemma. Although this aspect of Coase’s work is often overlooked or neglected in the law and economics literature, we believe it is Coase’s most original and counterintuitive insight. Consider, for

⁴⁶ Coase, *supra* note 5, at 2–15.

⁴⁷ *Id.*

⁴⁸ See Coase, *supra* note 5, at 2–15 (noting when “Coasean bargaining” will produce an efficient economic outcome). See also Ronald H. Coase, *The Federal Communications Commission*, 2 J.L. & ECON. 1, 25–30 (1959) (noting when “Coasean bargaining” will produce an efficient economic outcome).

⁴⁹ See STEVEN G. MEDEMA & RICHARD O. ZERBE, *The Coase Theorem*, 1 THE ENCYCLOPEDIA OF LAW AND ECONOMICS: THE HISTORY AND METHODOLOGY OF LAW AND ECONOMICS 837–38 (Boudewijn Bouckaert & Gerrit De Geest eds., 2000) (providing an extensive listing of some of the various statements of the Coase Theorem).

⁵⁰ Coase, *supra* note 5, at 2–15.

⁵¹ *Id.* at 1–2.

example, Coase's parable of the rancher and the farmer.⁵² According to Coase, it is a fallacy to think that the problem of cattle trespass is caused solely by the rancher.⁵³ In reality, cattle trespass (i.e., the risk of potential harm to the farmer's crops) is a joint problem.⁵⁴ Just as the rancher can reduce the risk of harm by reducing the size of his herd or erecting a boundary fence, so too can the farmer, either by planting cattle-resistant crops or by putting up the fence himself.⁵⁵ Likewise, the Prisoner's Dilemma also presents a reciprocal problem insofar as the payoffs for both prisoners stem from their independently made choices to defect or cooperate.⁵⁶ Thus, if both prisoners end up defecting in the standard one-shot version of the game (as game theory predicts they will do), then the prisoners have only themselves to blame for their collective plight.⁵⁷ In short, the prisoners' plight, like the problem of cattle trespass, is the product of a joint interaction: the outcome in both cases is not determined by the actions of just one party, but rather by the choices made by both of them jointly.

2. Well-Defined Property Rights

Does the second condition of the Coase Theorem (i.e., the legal assignment of well-defined property rights to one of the conflicting parties) apply to the Prisoner's Dilemma? If so, what property rights are being traded in the standard version of the Prisoner's Dilemma?

Recall that the Prisoner's Dilemma is a compelling parable about betrayal and loyalty, a secular morality tale about the conflict between individual and collective rationality. Strictly speaking, the Prisoner's Dilemma is not a story about property per se; however, property rights do play a secondary role in the dilemma. From a libertarian or classical liberal perspective, the prisoners have a vested property right in their personal liberty, and although personal liberty is often considered to be an *inalienable* (i.e., non-negotiable) right,⁵⁸ what is a plea bargain but

⁵² *Id.* at 2–15.

⁵³ *Id.*

⁵⁴ *Id.* at 1–2.

⁵⁵ See F. E. Guerra-Pujol & Orlando I. Martinez-Garcia, *Clones and the Coase Theorem*, 2 JOURNAL OF LAW & SOCIAL DEVIANCE 43, 65-73 (2011) (providing an extended discussion of the reciprocal nature of the rancher-farmer parable).

⁵⁶ Tucker, *supra* note 3.

⁵⁷ Tucker, *supra* note 3.

⁵⁸ See, e.g., Margaret Jane Radin, *Market Inalienability*, 100 HARV. L. REV. 1849, 1903–06 (1987) (discussing the commoditization of negative liberty). Of

a judicially sanctioned trade of one's personal liberty? When a criminal suspect is offered a plea bargain, the prosecutor is, in effect, asking the suspect to relinquish some of his personal liberty (i.e., he agrees to a certain but reduced prison sentence – x years) in exchange for avoiding the possibility of a maximum prison sentence (e.g., $5x$ or $10x$ years).⁵⁹

This broad definition of property (i.e., “liberty as property”) is consistent with traditional conceptions of property rights. The legal philosopher Stephen Munzer and the late political theorist C.B. Macpherson, among others, have described in detail different conceptions of property rights; in particular, property in the classical or common law sense refers to everything (tangible or intangible) to which a person has a right, including the right to personal liberty.⁶⁰ In the words of Macpherson, “men were said to have a property not only in land and goods and in claims on revenues for leases, mortgages, patents, monopolies and so on, but also a property in their lives and persons.”⁶¹ Although this classical conception of property is circular,⁶² our larger point here is that personal liberty is an intangible property right, a right that can be bargained away in certain situations, as in the Prisoner's Dilemma.

3. Zero Transaction Costs, Strategic Behavior, and Non-Strategic Bargaining

Stated in Coasean terms, the rules in the standard version of the Prisoner's Dilemma (i.e., no bargaining) artificially generate high transaction costs. But what if we change these rules to allow bargaining? That is, what if we imagine a Coasean version of the Prisoner's Dilemma, one with zero transactions costs?

Some scholars of the Coase Theorem, however, have already noted that parties, even parties who find themselves in a low

course, the most famous statement of this idea appears in the U.S. Declaration of Independence of July 4, 1776.

⁵⁹ Note that we do not mean to express our normative approval of plea bargains in criminal cases. We are simply making a descriptive point here about the secondary role of property rights in the Prisoner's Dilemma.

⁶⁰ See C. B. Macpherson, *The Meaning of Property*, PROPERTY: MAINSTREAM AND CRITICAL CONCEPTIONS 1, 8 (C. B. Macpherson, ed., Univ. of Toronto Press 1978) (identifying property as “a right – a somewhat uncertain right that has constantly to be re-asserted”); STEPHEN R. MUNZER, A THEORY OF PROPERTY 90 (Cambridge Univ. Press 1990) (identifying “liberty” among a list of items that should be considered personal goods (i.e. personal property) insofar as “they are often valued either in themselves or as means to other things that are valued or both”); Cheryl L. Harris, *Whiteness as Property*, 106 HARV. L.R. 1707, 1724–31 (1993) (providing a general overview of the broad historical concept of property).

⁶¹ C.B. Macpherson, *supra* note 64, at 7.

⁶² This conception of property is circular, since all it is saying, in effect, is that *one has a right to what one has a right to*.

transaction cost setting, may act strategically and thus fail to strike a mutually beneficial Coasean bargain.⁶³ Here, we shall contribute to the literature on the Coase Theorem in two ways. First, building on the work of Wayne Eastman, we shall consider strategic behavior in the form of promises and threats and explain why such strategic behavior might prevent the formation of Coasean bargains between the prisoners—even when they are allowed to communicate with each other. Next, building on the work of John Nash, we shall consider the possibility of non-strategic bargaining by the prisoners.

4. Strategic Bargaining, Threats and Promises in the Prisoner's Dilemma

One of the central lessons of game theory is that one can often – but not always, as we shall soon see – gain a tactical advantage during negotiations by committing oneself (or pre-committing, so to speak) to a particular strategy, such as a costly threat or an enforceable promise.⁶⁴ This insight is often referred to as the “first-mover advantage,”⁶⁵ and the ability to make one's threats or promises believable is considered a “credible commitment.”⁶⁶ The Coasean version of the Prisoner's Dilemma, however, poses an especially difficult challenge to the Coase Theorem because there is no first mover advantage in the Prisoner's Dilemma.

In summary, there is no first-mover advantage in the Coasean version of the Prisoner's Dilemma due to the possibility of strategic behavior. Assume, for example, that Prisoner 1 decides to

⁶³ See, e.g., Herbert Hovenkamp, *Marginal Utility and the Coase Theorem*, 75 CORNELL L. REV. 783, 787–91 (1990) (arguing that the failure of the Coase theorem to predict real world outcomes” is frequently explained by “the failure of the relevant actors” as opposed to “high transaction costs” Id. at 788); ROBERT D. COOTER & THOMAS ULEN, *LAW AND ECONOMICS* 242–44 (Scott Foresman & Co., 2d ed. 1982); and Robert Ellickson, “Of Coase and Cattle: Dispute Resolution Among Neighbors in Shasta County, 38 STAN. L. REV. 623, 625 n.4 (1986) (proposing that “negotiations in bilateral monopoly situations can be costly because the parties may act strategically”).

⁶⁴ See, e.g., AVINASH K. DIXIT & BARRY J. NALEBUFF, *THINKING STRATEGICALLY* 124–26 (W.W. Norton, reprinted ed. 1993) (asserting that strategic moves are two-pronged: (i) the planned course of action and (ii) the commitment that makes this course credible); see also DOUGLAS BAIRD, *ET AL.*, *GAME THEORY AND THE LAW* 43-44 (1994).

⁶⁵ Roger A. Kerin, P. Rajan Varadarajan & Robert A. Peterson, *First-Mover Advantage: A Synthesis, Conceptual Framework, and Research Propositions*, 56 J. MARKETING, Oct., 1992, at 33, 33.

⁶⁶ See, e.g., Douglass C. North, *Institutions and Credible Commitment*, 149 J. INST'L THEORETICAL ECON 11-12 (Mar. 1993) (identifying that the enforcement is “central to credible commitment”).

pre-commit to cooperation and is able to communicate his cooperative commitment to Prisoner 2, say by taking a sincere and solemn oath in Prisoner 2's presence to remain silent "no matter what."⁶⁷ Also, assume that Prisoner 2 truly believes in the sincerity and seriousness of the other prisoner's solemn oath. That is, he knows that Prisoner 1 is a man of his word. Perversely, the logic of defection continues to prevail, for defection or confessing is still Prisoner 2's dominant strategy. In fact, Prisoner 2 has an even stronger incentive to defect in this situation because he is now certain to obtain the temptation payoff, given the other prisoner's binding promise not to defect.

Knowing this, what if Prisoner 1 took a different approach and made a credible threat instead of a mere promise? That is, assume now that Prisoner 1 is able to make and communicate a credible threat to punish the other prisoner in the event that the latter decides to defect. Introducing the tactical use of a credible threat, however, dramatically changes the payoffs of the game.⁶⁸ In other words, an enforceable agreement backed up by a credible threat changes the values of the payoffs of the prisoners.⁶⁹ Therefore, strictly speaking, under these facts, the prisoners are no longer playing a Prisoner's Dilemma.

Stated formally, a credible threat changes Prisoner 2's temptation payoff, T ; specifically, the value of T decreases as the severity of the threat increases.⁷⁰ But, let us put this technical objection to one side and consider the possibility of non-strategic bargaining by the prisoners in the Coasean version of the dilemma.

5. Non-Strategic Coasean Bargaining

Assume that the prisoners can bargain with each other and can make credible threats and binding promises.⁷¹ Without a

⁶⁷ For clarity, we shall follow Luce & Raiffa's interpretation of the parable and continue to refer to Player A as "Prisoner 1" and Player B as "Prisoner 2" in the remainder of this paper.

⁶⁸ Cf. Wayne Eastman, *Everything is up for Grabs: The Coasean Story in Game-Theoretic Terms*, 31 NEW ENG. L. REV. 1, 1-37 (1996) (discussing the idea of "payoff mutability")

⁶⁹ See, e.g., Elinor Ostrom, *et al.*, *Covenants with and without a Sword: Self-Governance Is Possible*, 86 AMERICAN POLITICAL SCIENCE REVIEW 404, 413-414 (1992) (reviewing the "payoff consequences" of selecting or not selection a sanctioning mechanism in a common-pool resource game).

⁷⁰ Recall that a Prisoner's Dilemma occurs when the values of the payoffs are $T > R > P > S$. (See *supra* part I.B.) The employment of a credible threat, however, changes this payoff structure to $R > P > S > T$, or to $R > P > T > S$, or perhaps to $R > T > P > S$, depending on the severity of the threat and the resulting new value of T .

⁷¹ Recall that the payoffs in the dilemma can be stated numerically or algebraically by the variables T , R , P , and S . We will follow this convention in the remainder of this paper.

Coasean bargain, both prisoners will most likely end up confessing – or “defecting” in the parlance of game theory – because defection is the only Nash equilibrium in the standard one-shot version of the Prisoner’s Dilemma. Since the defection payoff is equal to P (i.e., the “punishment” payoff for mutual defection), Prisoner 1’s payoff is equal to P_1 , while Prisoner 2’s payoff is P_2 . In the standard version of the Prisoner’s Dilemma, however, P_1 is equal to P_2 , since the payoffs are symmetrical. Following convention, and because these are the payoffs the prisoners will most likely receive if they are unable to bargain with each other, we shall refer to these defection payoffs as the “outside options” or “disagreement values” of the prisoners.⁷²

If, however, the prisoners agree to cooperate – a likely outcome if bargaining is allowed – the prisoners will receive R , the “reward” payoff for mutual cooperation. Therefore, both prisoners are better off cooperating because cooperation produces a collective gain for both prisoners (i.e., $R > P$). Or put another way, the gains from a Coasean bargain in the Prisoner’s Dilemma are positive (i.e., $R - P > 0$). (This all assumes, of course, that neither prisoner breaches the agreement – a possibility that we will explore later.)

But how will the prisoners split the collective gains from their Coasean bargain? Stated formally, Prisoner 1 will receive $(R + P_1 - P_2)/2$, and Prisoner 2 will receive $(R + P_2 - P_1)/2$. Therefore, each prisoner’s share of the payoffs depends, not only on the value of his gains from trade or the Coasean bargain (i.e., the reward payoff, R), but also on the prisoners’ outside options or disagreement values (i.e., P_1 and P_2).⁷³ Nevertheless, in the standard version of the dilemma, since the prisoners’ outside options are the same (i.e., $P_1 = P_2$) neither prisoner in the Coasean version of the game can improve his bargaining position by improving his outside option or decreasing that of the other prisoner (i.e., each prisoner’s payoff for mutual cooperation is equal to $R/2$). Accordingly, since the payoffs in the standard version of the Prisoner’s Dilemma are symmetrical, they will split their gains evenly.

Thus far, this analysis suggests that the prisoners have every incentive to strike a Coasean bargain and cooperate, so long as $R/2 > P$. But, notice what this analysis does *not* tell us. It does not tell us whether the prisoners will, in fact, keep their mutual

⁷² Cf. LUKE M. FROEB, BRIAN T. MCCANN, MIKHAEL SHOR & MICHAEL R. WARD, *MANAGERIAL ECONOMICS: A PROBLEM SOLVING APPROACH* 190 (3d ed., 2014).

⁷³ See John F. Nash, *The Bargaining Problem*, 18 *ECONOMETRICA* 155, 157-158 (1951). Note that Nash uses the term “anticipations” to refer to the outside options or disagreement values of the players.

promises or whether they will breach them. In fact, a Coasean bargain may not solve the Coasean version of the Prisoner's Dilemma because T , the temptation payoff, still lurks in the background. So long as T remains larger than R , each prisoner has a countervailing incentive to breach his promise of cooperation: the larger T is, relative to R , the more likely it is that one or both of the prisoners will defect.

IV. THE ROLE OF UNCERTAINTY, EXPONENTIAL DISCOUNTING, AND ELASTICITY IN THE COASEAN VERSION OF THE PRISONER'S DILEMMA

Even when the prisoners are allowed to bargain with each other – either strategically or non-strategically, as in our Coasean thought-experiment above – and even when they are able to make credible threats, the prisoners may still end up defecting. It is true that the use of credible threats might change the temptation payoff relative to the other payoffs; however, there are three non-trivial reasons why it might not. First, uncertainty poses a major problem with threats, since there will always exist some level of uncertainty as to whether a threat will in fact be carried out. Another salient problem with threats is *exponential discounting* or the time dimension of a given threat; this is particularly relevant since most threats, however credible, will not be carried out until sometime in the future. Lastly, another potential problem with threats is the issue of *price elasticity of demand*, since the prisoners' responsiveness to a threat may vary depending on a number of factors.

A. *Uncertainty*

"Uncertainty" refers to the positive probability that any Coasean bargain made between the prisoners will not be enforced due to judicial error or some other extrajudicial factor. Essentially, one's decision to defect in the Coasean version of the dilemma will not only be a function of the severity of the penalty for breach (i.e., any threats or promises made during the course of the prisoners' negotiations), but it will also be a function of the probability of enforcement. Both of these functions – severity of penalty and probability of enforcement – are uncertain *ex ante* (i.e., at the time one must decide whether to defect or not). Generally speaking, the less likely enforcement is, or the less severe the penalty for breach is, the more likely the prisoners will defect.

One possible response to the problem of uncertainty is to extend the logic of zero transactions costs to the enforcement stage. Since the Coase Theorem assumes costless bargaining, why not further assume costless enforcement? Could we not assume that the prisoners are not only allowed to bargain and make credible threats and promises, but also that any resulting

agreement to cooperate will be enforced perfectly and costlessly? This, in turn, raises a new question: does Coasean bargaining solve the Prisoner's Dilemma even when enforcement is costless and perfect? Not necessarily, for the answer to our question now depends on how far in the future such enforcement will occur.

B. Exponential Discounting

The next question we shall consider is what role does time play in the Coasean version of the Prisoner's Dilemma? In general, notice that the Prisoner's Dilemma presents an intertemporal choice. Each prisoner must weigh not only the probability that the other will defect in the absence of a Coasean bargain (or the probability of breach even with a Coasean bargain), but each prisoner must also weigh the present value of his own defection or breach versus the future value of cooperation.

Assume that the prisoners are allowed to bargain with each other and have each promised to cooperate in order to obtain the higher collective payoffs generated from mutual cooperation. Even with a Coasean bargain in place, each prisoner must weigh the present value of breaching his promise (i.e., defecting) versus the future or discounted value of cooperating (i.e., keeping his promise). That is, each prisoner must still decide whether he prefers a reduced sentence in the present, which is a higher payoff relative to his other choices, versus the possibility of a penalty for breach in the future.

According to the standard economic model of behavior, intertemporal choices are no different from other choices, except that some consequences are delayed and hence must be anticipated and "discounted" (i.e., recalibrated to take into account the delay).⁷⁴ But discounting generates the possibility of "exponential discounting." That is, given two similar rewards, people generally prefer the one that arrives sooner rather than the equivalent one later. Stated formally, people often "discount" or reduce the value of the later reward by a factor that increases with the length of the delay. This discounting process is traditionally modeled in economics as a form of exponential discounting, a time-consistent model of discounting.⁷⁵

⁷⁴ See, e.g., A DICTIONARY OF ECONOMICS 108–09 (John Black, Nigar Hashimzade & Gareth Myles eds., 4th ed. 2012) (providing a standard definition of "discount" and "discounting the future" in economics).

⁷⁵ Some experimental research has shown that the constant discount rate assumed in exponential discounting is systematically being violated. Shane Frederick, George Loewenstein & Ted O'Donoghue, *Time Discounting and*

In the case of the Prisoner's Dilemma, our prisoners are more likely to engage in exponential discounting when calculating the present value of a reduced prison sentence (i.e., the greater their discount rate is, the more they value present personal liberty over future liberty, and thus the more they value present liberty over future liberty, the more likely they are to defect). Does a Coasean bargain between the prisoners change this outcome? Not at all – the outcome will not change if the present value or utility of a reduced sentence today is greater than the expected or discounted disutility of a penalty for breach in the distant future. In other words, it is possible that the temptation payoff, which is certain and will occur at time T1, might outweigh the possibility of a breach penalty, which is uncertain and will not occur until time T2.

Thus, because any penalty for breach will occur in the future, the present utility from a (certain) reduced prison sentence now is likely to outweigh the future disutility of an (uncertain) penalty for breach in the future! Of course, whether the discounted disutility of a future penalty for breach outweighs the present value of a reduced sentence depends on several critical variables, including (i) the size of the future or expected penalty, (ii) the probability that the breach is enforced, and (iii) each prisoner's discount rate. More to the point, however, we have identified the conditions under which our prisoners are likely to defect even with a Coasean bargain in place. And, even under the standard assumptions in modern economic theory, these conditions are not implausible or far-fetched.

Compare, for example, the related idea of interest (i.e., time value of money), a foundational concept in finance theory.⁷⁶ A certain amount of money today has a different buying power (value) than the same amount of money in the future because the value of money at a future point of time includes the interest earned or inflation accrued over a given period of time.⁷⁷ In the alternative, the time value of money can also be stated formally: the sum of *FV* (future value) to be received in one year is discounted at the rate of interest *r* to give a sum of *PV* (present value) at present (i.e., $PV = FV - r*PV = FV/(1+r)$). This expression measures the present value of a future sum, discounted to the present by an amount equal to the time value of money. In other words, this concept allows the valuation of a future stream of

Time Preference: A Critical Review, 40 J. ECON. LITERATURE 351 (2002). This paper, however, will follow the standard economic approach and assume a constant discount rate.

⁷⁶ DAVID G. LUENBERGER, INVESTMENT SCIENCE ch. 2 (Oxford Univ. Press 1998).

⁷⁷ That is, the value of money changes over time because there is an opportunity to earn interest on the money and because inflation will tend to drive prices up, thus reducing the "value" of the money in the future. *Id.* at 12.

income, such that the future stream is “discounted” and then added together, thus providing a lump-sum “present value” today of the entire income stream.

Like the time value of money, there is also a *time value of time*, so to speak. One way of measuring the magnitude of each prisoner's incentive to breach (i.e., the probability that either prisoner will breach or defect), even with a Coasean bargain in place, is by analyzing the role that time plays in his or her decision-making. The prisoners not only prefer personal liberty to the absence thereof (time in prison), but we would also expect the value or utility of liberty in the present to be worth more to each prisoner than liberty in the future. In other words, the “time value of time” means that personal liberty in the present is worth more than in the future, and likewise, time in prison in the present imposes a greater disutility than time in prison in the future. In addition, independent of the effect that time has on the decision-making of the prisoners, we must further consider the prisoners' responsiveness to the payoffs in the Prisoner's Dilemma. That is, in predicting whether the prisoners will defect or cooperate in the Coasean version of the dilemma, the price elasticity of demand of each prisoner must also be considered.

C. Price Elasticity of Demand

Here, we pose one last important question regarding our Coasean version of the Prisoner's Dilemma. If a prison sentence operates like a price,⁷⁸ then what happens when the price elasticity of demand of each prisoner is different? In economics, the term “elasticity” generally refers to the percentage change in one variable with respect to a percentage change in another variable, or the ratio of the logarithmic derivatives of the two variables.⁷⁹ Specifically, the “price elasticity of demand” is a numerical or quantitative measure of how responsive the demand of a given good or service is to a change in the price of that good or service. In the case of the Prisoner's Dilemma, for example, the “good” being demanded by the prisoners is personal liberty (i.e., a reduced prison sentence). Elasticity in this case would thus measure the prisoners' responsiveness to changes in the prison sentence.

⁷⁸ See generally Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 JOURNAL OF POLITICAL ECONOMY 169, 179-180 (1968).

⁷⁹ HARRY H. PANJER, FINANCIAL ECONOMICS WITH APPLICATIONS 101 (Actuarial Found. 1998). See also CAMPBELL MCCONNELL, STANLEY BRUE, AND SEAN FLYNN, ECONOMICS 114-116 (18th ed. 2009).

Thus far, however, we have assumed that the prisoners' elasticities are the same, a standard but unstated assumption in most, if not all, treatments of the Prisoner's Dilemma. Specifically, we have assumed that both prisoners share a "unitary elastic" (i.e., $e = 1$) demand schedule. In other words, we have assumed that the prisoners share the same set of time preferences regarding the payoffs in the Prisoner's Dilemma: they both uniformly prefer shorter prison sentences to longer ones. Stated formally, we have not only assumed that the prisoners derive a greater amount of utility (or a lower level of disutility) the shorter their prison sentences are, but we have also assumed that the prisoners obtain the same levels of "utility" or "disutility," as the case may be, from the payoffs (prison sentences) in the Prisoner's Dilemma. (Note that in economics, "utility" is an abstract or mathematical representation of preferences over some set of goods and services.⁸⁰) In the case of the Prisoner's Dilemma, an additional unit of time in prison generates an additional, and perhaps diminishing, level of *disutility* on the prisoners.

Thus, the question above (i.e., what happens if the prisoners' elasticities of demand are different?) becomes: what happens when Prisoner 1's demand for personal liberty (i.e. a reduced sentence) is inelastic (i.e., $e_1 < 1$), while Prisoner 2's demand for liberty is elastic (i.e., $e_2 > 1$)?

Before we proceed to answer this question, let us explain "inelastic" and "elastic" demand and illustrate these concepts with a simple numerical example. The demand of a good is "elastic" (i.e., more responsive to price changes) when the percentage change in the price of that good is less than the percentage change in quantity demanded.⁸¹ For example, when $e = 1.5$, this means that a 50% decline in price will cause a 75% increase in the quantity demanded.⁸² In contrast, demand is "inelastic," or less responsive to changes in price, when the percentage change in the price of a good exceeds the percentage change in quantity demanded.⁸³ For example, when $e = 0.5$, this means that a 50% decline in price will only cause a 25% increase in the quantity demanded.⁸⁴ Thus, in the case of the Prisoner's Dilemma, the concept of elasticity refers to the prisoners' responsiveness to changes in the payoffs of the game. For example, Prisoner 1 might be highly responsive to small changes in the prison sentence; as such, his demand for personal liberty would be elastic. On the

⁸⁰ For an influential treatment of utilities in economics, see VON NEUMANN & MORGENTHAU, *supra* note 21, at ch. 3, 17–31 (providing a mathematical treatment of utilities and assigning utilities to probability distributions of alternatives).

⁸¹ COOTER & ULEN, *supra* note 67, at 29.

⁸² *Id.*

⁸³ *Id.*

⁸⁴ *Id.*

other hand, Prisoner 2 might be far less responsive even to large changes in the payoffs, and therefore, his demand for liberty would be inelastic.

The most important determinant of the price elasticity of demand is the availability of substitutes for the good in question.⁸⁵ Generally speaking, the elasticity of demand will be greater where there are more substitutes for a particular good, and, likewise, the elasticity will be lower where there are fewer substitutes.⁸⁶ In the case of the Prisoner's Dilemma, however, the responsiveness of the prisoners to the payoffs may vary depending on certain individual factors unique to each prisoner. Although there are few, if any, substitutes for personal liberty (i.e., you are either free or in prison), the level of disutility of being in prison may vary depending on a wide variety of individual factors, such as, *inter alia*, one's age, income, marital status, or history of prior convictions. We would expect a young prisoner, a wealthy prisoner, or a prisoner with a wife and children, for example, to behave differently than an old prisoner, a poor one, or one with no family ties. Likewise, a prisoner who is a first-time offender, might qualify for probation or a rehabilitation program, whereas a repeat offender might face a mandatory-minimum prison term. In addition, we would expect the *quality* of the prison sentence or type of prison (i.e., a high security prison with limited visitation rights versus a low security, college-campus type prison with a good library, internet access, and liberal visitation rights) – and not just the *quantity* of time in prison – to influence the behavior of the prisoners in the Prisoner's Dilemma.

In other words, the use of general labels, such as "Prisoner 1" and "Prisoner 2" (or "A" and "B"), to describe the players in the Prisoner's Dilemma is too reductionist and possibly even misleading because such labels abstract away the problem of elasticity. Accordingly, we need more – not less – information about the prisoners' individual circumstances and specific characteristics in order to measure their respective responsiveness to the payoffs in the Prisoner's Dilemma game.

Instead of ignoring this critical information, let us now proceed under a different set of assumptions. In the following three examples, assume that we have sufficient information about the individual prisoners in order to measure or at least approximate their actual elasticities. Example #1 assumes that the price elasticity of demand for personal liberty of both prisoners

⁸⁵ *Id.* at 29.

⁸⁶ "The more substitutes for a good, the greater the elasticity of demand; the fewer the substitutes, the lower the elasticity." *Id.* at 29–30.

is elastic (i.e., $e > 1$). Example #2 considers the more interesting case of a highly inelastic prisoner playing against a highly elastic one. And, Example #3 considers prisoners with inelastic demand curves (i.e., $e < 1$).

Example #1: $e > 1$

To begin, assume that both prisoners are highly elastic (i.e., responsive) to changes in the payoffs in the standard version of the Prisoner's Dilemma. In this case, we would expect no change in the prisoners' responses to the payoffs in the game because their levels of utility or disutility from the payoffs remain unchanged relative to each other. So long as the responsiveness of the prisoners to changes in the payoffs run in the same direction (i.e., so long as both prisoners are price elastic or price inelastic), both prisoners still prefer to spend less time in prison to more time.

Example #2: $e > 1, e < 1$

Next, consider the more interesting case of a highly *inelastic* prisoner playing against a highly *elastic* one. Contrary to the first example, assume that the corresponding elasticities of the prisoners in the standard one-shot version of the Prisoner's Dilemma run in opposite directions: Prisoner 1's demand for personal liberty is highly *elastic* (i.e., $e_1 > 10$), while Prisoner 2's desire to stay out of prison is highly *inelastic* (i.e., $e_2 < 0.1$). Under these conditions, both prisoners still prefer short prison sentences to long ones, but Prisoner 1 is much more responsive to any changes in the payoffs of the Prisoner's Dilemma than Prisoner 2 is. Does this scenario alter the likely outcome or equilibrium of the dilemma?

We believe it does. Under this scenario, Prisoner 1 is much more likely to defect than Prisoner 2 because Prisoner 1, as "defined" by his elasticity curve, is more responsive to the payoffs of the game. In particular, Prisoner 1 – like Prisoner 2 – wants (i) the lowest possible sentence (i.e., T , the Temptation Payoff) and (ii) to avoid the worst possible payoff (i.e., S , the dreaded Sucker's Payoff). However, Prisoner 1 – *unlike* Prisoner 2 – is more responsive to the possibility of (i) obtaining the Temptation Payoff, as well as (ii) avoiding the humiliating Sucker's Payoff.

What about Prisoner 1's inelastic cohort, Prisoner 2? By definition, Prisoner 2 is less responsive to changes in the payoffs than Prisoner 1 because Prisoner 2's demand for liberty is highly *inelastic* (i.e., $e_2 < 0.1$). Prisoner 2's behavior, therefore, will be much harder to predict for multiple reasons. On the one hand, Prisoner 2 – like all prisoners, presumably – prefers a short prison sentence to a long one. On the other hand, Prisoner 2 (i.e., $e_2 < 0.1$) is less responsive to changes in the payoffs than the average

prisoner (i.e., $e = 1$), and is far less responsive to such changes than Prisoner 1 (i.e., $e_1 > 10$). We would thus expect Prisoner 2 to be highly unresponsive to the prosecutor's strategic offer of leniency in exchange for his confession.

Therefore, whether Prisoner 2 decides to defect or to cooperate will, most likely, depend on his personal value system and other relevant or applicable extra-strategic factors (e.g. age, income, marital status, etc.). And yet, it is these factors that are completely ignored or abstracted away in game theory. Put another way, if Prisoner 2 is already predisposed to reject any potential plea bargain or offer of leniency from the prosecutor (e.g. because of Prisoner 2's value system), then he is unlikely to confess or accept a plea bargain *ex post* (i.e., *after* the prosecutor's offer of a reduced sentence is on the table).⁸⁷

Example #3: $e < 1$

Lastly, what happens when both prisoners' demand curves are highly inelastic? Or, what is the most likely outcome or equilibrium of the game when both prisoners are highly unresponsive to changes in the payoff structure of the Prisoner's Dilemma? Simply put, all bets are off in this scenario. Similar to the discussion concerning Prisoner 2 in example #2 above, factors external to the Prisoner's Dilemma model will influence the behavior of the prisoners in this example more than the actual payoffs.

D. Lessons and Discussion

These three examples of the role of elasticity in the Prisoner's Dilemma teach us an important and non-trivial lesson about the Prisoner's Dilemma model and about game theory in general. Game theory is best able to predict the behavior of players in the Prisoner's Dilemma (and other games) when their demand curves are inelastic (i.e., $e < 1$) but not when the demand schedules are elastic (i.e., $e > 1$) or when their elasticities are unitary (i.e., $e = 1$). Since the behavior of such inelastic players will depend less on the payoffs of a given model and more on real-world factors outside of the formal model, the predictive power of game theory will decrease as the prisoners' preferences become more responsive (i.e., their demand curves become more elastic). Indeed, this lesson

⁸⁷ But it is worth noting that if Prisoner 2 is already predisposed *ex ante* to confess or strike a deal with the prosecutor (for reasons not captured in the abstract Prisoner's Dilemma model), then he will probably still confess *ex post*, despite his highly inelastic demand curve.

is not only consistent with one of the key insights of Thomas Schelling's classic study "The Strategy of Conflict."⁸⁸ It also builds upon Schelling's seminal work by specifying the limits of game theory. By studying the theoretical relation between the behavior and choices of the players and their respective elasticities of demand, our work has identified circumstances in which game theory models are likely to be helpful and when they are likely to prove incomplete, misleading, or wrong.

V. A BRIEF DIGRESSION REGARDING THE ROLE OF THIRD PARTIES IN THE PRISONER'S DILEMMA

Before proceeding any further, we shall return one last time to the standard, or non-Coasean version, of the Prisoner's Dilemma to explore the relation between the prisoners and the prosecutor in the standard version of this parable. Stated in general terms, we shall consider the relation of the "third-party payoff administrator" to Players 1 and 2 in the general or logical form of the game.

Whatever one thinks of our Coasean thought-experiment or Coasean version of the Prisoner's Dilemma, it is worth noting that Coasean bargaining is already taking place, even in the standard versions of the parable presented above. But instead of direct bargaining between the prisoners themselves (which as we saw is not allowed in the standard version of the Prisoner's Dilemma), the bargaining that is taking place in this game is between each prisoner and the prosecutor separately.

The standard formulations of the Prisoner's Dilemma presuppose not just two prisoners or players but also a "third-party payoff administrator" (such as the prosecutor in the original formulation of the parable). That is, in addition to the players or prisoners, the Prisoner's Dilemma also requires a third-party to administer the payoffs of this game, with payoffs depending on the choices made by the players.⁸⁹ This third party is not really a neutral arbiter or mere "payoff administrator." Instead, he is trying to manipulate the choices of the players by getting them to confess or "snitch" in the classic version of the parable, and, moreover, his conduct is another form of "bargaining" with the players.

The presence of the prosecutor or "third-party payoff administrator" in the standard versions of the parable thus poses an important but neglected subsidiary question.⁹⁰ Doesn't the

⁸⁸ See generally THOMAS C. SCHELLING, *THE STRATEGY OF CONFLICT* (2d ed. 1980).

⁸⁹ For example, Richard Dawkins refers to the role of the "banker" in his presentation of the parable. RICHARD DAWKINS, *THE SELFISH GAME* 203, 206–07, 217–18, 225–26 (2d ed. 1989).

⁹⁰ See, e.g., F. E. Guerra-Pujol, *The Poker-Litigation Game* 3, n.5 (Dec. 26,

presence of this third party (i.e., his ability to offer lighter sentences or more favorable payoffs to the prisoners) affect the outcome of the game? Would the prisoners still defect in the one-shot version of the parable if the role of the prosecutor or other third party were removed from the game?

Recall that the standard or “canonical” version of the Prisoner’s Dilemma is classified as a “non-cooperative game” because the prisoners in the dilemma are not allowed to communicate or negotiate *with each other*.⁹¹ Nevertheless, although the prisoners are not allowed to bargain with one another, it is critical to note that the prosecutor is, in fact, allowed to communicate and bargain with the prisoners. The prosecutor in the standard versions of the dilemma is, in essence, bargaining with each prisoner separately and sequentially, making a tempting “take it or leave it” offer to each one. Although neither prisoner is allowed to make a counteroffer to the prosecutor, each prisoner must still decide whether to accept the prosecutor’s initial offer. In the standard one-shot version of the dilemma, both prisoners will most likely accept the prosecutor’s offer (i.e., agree to confess), because confession is the dominant strategy or Nash equilibrium of this game.

In short, the prisoners are, in fact, already engaged in a form of Coasean or voluntary bargaining in the standard version of the Prisoner’s Dilemma. Although they are not allowed to bargain with each other, they are allowed to bargain, so to speak, with the prosecutor. But, the collective outcome of these separate Coasean bargains with the prosecutor leaves both prisoners much worse off than if they had decided to reject the prosecutor’s offer and remain silent.

This analysis of the dilemma thus refutes the Coase Theorem; it shows how self-seeking Coasean bargaining (i.e., Coasean bargaining between each prisoner and the prosecutor) generates a worse collective outcome for the prisoners. One could argue that this conclusion is premature because, given the structure of the payoffs in the standard version of the Prisoner’s Dilemma, it is very likely that the prisoners would have defected anyways. But this conclusion is not premature at all, at least not with respect to the Prisoner’s Dilemma. For the prisoners to defect, they must be able to strike a bargain with the prosecutor. That is, there must be

2012) (unpublished manuscript) (on file with author), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2193993 (providing a simple model in which the role of the banker is made explicit).

⁹¹ See Tucker, *supra* note 3 and accompanying text; LUCE & RAIFFA, *supra* note 17, at 94–95 and accompanying text.

someone (i.e., the prosecutor) with the ability to offer a lighter prison sentence in exchange for the prisoners' confessions. By contrast, if the District Attorney is prevented from bargaining with the prisoners, or if the prisoners are prevented from bargaining with the D.A., then it is less likely that the prisoners will end up defecting. In short, in a world in which "plea bargaining" is prohibited, the prisoners are probably better off going to trial and taking their chances.

Despite this analysis, most game theorists would probably agree that, due to the structure of the payoffs in the standard "one-shot" version of the Prisoner's Dilemma, defection is still the most likely outcome in one-shot dilemmas – even when all bargaining is prohibited. Once we allow Coasean bargaining between the prisoners, however, we see that there are three sets of potential bargains in the Prisoner's Dilemma. Specifically, there is the possibility of a Coasean bargain between the prisoners themselves, especially in the Coasean version of the dilemma, but there is also the possibility of a separate bargain between Prisoner 1 and the prosecutor as well as the possibility of an additional bargain between Prisoner 2 and the prosecutor. The possibility of three separate sets of bargains in the Prisoner's Dilemma suggests that the outcome of such a three-person interaction might be a complex one and possibly unpredictable. We thus conclude this paper by conducting a preliminary exploration of the relation between complexity theory and the Coasean version of the Prisoner's Dilemma.

VI. SOME CLOSING THOUGHTS ON THE COMPLEXITY OF THE PRISONER'S DILEMMA

The classic or standard version of the Prisoner's Dilemma paints a misleading picture of the game being played and the number of players. It purports to be a *two-player* model when, in reality, there are at least *three different persons* playing this game: the two prisoners as well as the prosecutor (and the police). Therefore, instead of a dyad or two-party interaction, we have a triad or three-party interaction, one that is more complex and with many more relevant variables. Such stories as the Prisoner's Dilemma and the Rancher-Farmer Parable, however, purposely ignore such endogenous and exogenous variables – variables that could very well influence the outcome of these interactions. In a real-life Prisoner's Dilemma, for example, the prisoners are likely to find themselves embedded in a larger network of players, all of whom are ignored in the existing legal and economics literature on the Coase Theorem and the Prisoner's Dilemma.

Suffice it to say, the different variables that shape the preferences of the prisoners and the prosecutor make the Prisoner's Dilemma a potentially very complex game. Moreover, as

more variables and degrees of elasticities influence the triadic relation among the prisoners and prosecutor, the more complex their interaction becomes. Such a triadic and multivariable interaction thus invites the use of a different approach, such as complexity theory.⁹² This, however, will be the subject of a future paper.

VII. CONCLUSION

Before concluding, we wish to say a few words about our general approach to the question posed in the title of our paper as well as our emphasis on *questions* (as opposed to *answers*) or “known unknowns”⁹³ throughout this paper. To paraphrase Stuart Firestein, a neurobiologist at Columbia University, our implicit premise in these pages is that communal ignorance (that which we do not yet know) is the main fountain of knowledge and discovery.⁹⁴ According to Firestein, ignorance promotes discovery because it motivates persons engaged in science to search for answers, and this pursuit, in turn, leads to new questions: “[ignorance] is not an individual lack of information but a communal gap in knowledge . . . This is knowledgeable ignorance, perceptive ignorance, insightful ignorance. It leads us to frame better questions, the first step to getting better answers.”⁹⁵ We believe this counterintuitive and critical logic also applies to economics and law, and to the social sciences generally. Rather than restating what we already know (or think we know), as many conventional legal scholars and economists tend to do, we make greater progress when we pose new and non-trivial questions (i.e., questions to which we do not yet know the answers).

In this paper, then, we identified the essential elements of the one-shot, two-player Prisoner's Dilemma, the simplest and most famous of all models in game theory, and then presented a pure Coasean version of the dilemma, one in which the prisoners are allowed to communicate and bargain with *each other*, and not just with the prosecutor. We found that even when the prisoners are allowed to communicate and bargain with each other, there is

⁹² See generally MELANIE MITCHELL, *COMPLEXITY: A GUIDED TOUR* (2009). For applications to law, see Orlando I. Martínez-García, *The Person in Law, the Number in Math*, 18 AM. U. J. OF GENDER SOC. POLY & L. 503 (2010).

⁹³ Moran Cerf, *Known Unknowns*, 336 SCI. 1382 (2012) (reviewing STUART FIRESTEIN, *IGNORANCE: HOW IT DRIVES SCIENCE* (2012)).

⁹⁴ By “ignorance,” we follow Firestein in meaning “the absence of fact, understanding, insight, or clarity about something.” STUART FIRESTEIN, *IGNORANCE: HOW IT DRIVES SCIENCE* 6 (2012).

⁹⁵ *Id.* at 7.

some positive probability that they might not strike a Coasean bargain. Furthermore, we found that even if they are able to negotiate a mutually beneficial agreement (e.g. through non-strategic bargaining), there is also some positive probability that they could still breach such an agreement and end up defecting, contrary to what the Coase Theorem predicts. In either case, the probability of defection is a function of various factors, including such things as uncertainty, exponential discounting, and elasticity.

This conclusion – the possibility of defection in the Coasean version of dilemma – is theoretically significant because it all but refutes or falsifies the Coase Theorem. It is also worth noting that our conclusion is not based on *ad hoc* behavioral or psychological quirks of human behavior. Uncertainty, exponential discounting, and elasticity are all part of the standard economics toolkit and are based on the standard rationality assumption of economics. The main contribution of the thought-experiment presented in this paper – our Coasean version of the Prisoner's Dilemma – is that it poses many deep and difficult questions, and this paper is our first attempt in search of answers ... and new questions.