AN ECONOMIC MODEL OF PLEA BARGAINING IN THE CRIMINAL COURT SYSTEM

Dissertation for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

JUDITH ANN LACHMAN

1975

This is to certify that the

thesis entitled

AN ECONOMIC MODEL OF
PLEA BARGAINING
IN THE CRIMINAL COURT SYSTEM

presented by

Judith Ann Lachman

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Economics

William J. Haley
Major professor

Date September 19, 1974

O-7639

ABSTRACT

AN ECONOMIC MODEL OF PLEA BARGAINING IN THE CRIMINAL COURT SYSTEM

By

Judith Ann Lachman

Plea bargaining, the informal process by which prosecution and defense settle a criminal case without trial, is the dominant method of disposing of criminal cases in U. S. metropolitan courts. Both the process and outcome of plea bargaining lend themselves to analysis within an economic framework: each party makes rational choices based upon its estimate of the opportunity cost of a guilty plea, that is, the expected outcome and associated costs of trial.

This dissertation presents a theory of plea bargaining--the derivation of the reservation prices of the prosecutor and defendant, the offer-response pattern, and the outcome of bargaining--and tests that theory using micro cross-sectional data.

In the theory, each of the parties to negotiations is believed to calculate the outcome and utility anticipated from trial, and then the minimal concession which would be required from his opponent in order to make the prospect of a plea of guilty equally attractive. For each

party, the trial option is represented by one constraint and the guilty plea option by another. The "switch function" which results from repeating the above calculation for alternative values of the independent variable is in effect the demand function for one good--a plea of guilty--stated in terms of the price of a substitute good, trial. Each opportunity set is a production possibility frontier derived from functions describing case outcome in terms of variable inputs to production of a strong case (additional sleuthing and research efforts) and the relative strength of the two parties. Concession is interpreted to be the deliberate underutilization of the strength of one's own case.

The integration of the above items within the model leads to two observable theoretical results. First, the derivation of the switch functions--each of which divides the strength-outcome space into a set of acceptable bargain outcomes and a set of unacceptable ones--lends a specific functional definition to the concept of a Pareto optimal set of outcomes of negotiation; the set of negotiated outcomes should fall within the region bounded by the switch functions. Due to the lack of statistical tools, this can only be "tested" by visual observation of data plots.

Second, a split-the-difference theory of bargained outcomes is proposed, and the outcome function that suggests is a linear combination of the two switch functions. Estimation of this outcome function serves not only to test the general validity of the model, but also to determine the proportions in which the bargainable difference is split.

In order to estimate outcome as a function of relative strength, it is necessary first to create an index of relative strength based upon the factors believed to be critical to the determination of strength at trial. This is done using probability-of-conviction estimates made by prosecutors together with information on characteristics of the defendant and of the offense, and the initial endowment of evidence. A logistic form is employed, and predicted values for the probability estimates are used in subsequent empirical work.

Regression results obtained for the relative strength indicator suggest that delay is beneficial to defendants, as is having a less harsh judge, or knowing the complainant well. In contrast, having other charges pending or being age 18 or older work to the defendant's disadvantage. The most significant results here relate to the race and jail status of the defendant, and interactions between these variables. Taken singly, the effects of each of race and jail on strength at trial are as expected: blacks fare worse than whites, and pretrial incarceration works against the defendant. However, the effect of race reverses for defendants in jail awaing trial. Whether this occurs because of racial discrimination in favor of blacks, or of class discrimination against higher income blacks, or of racial discrimination <u>against</u> blacks--by incarceration of "good risk" blacks while their white counterparts await trial at liberty--must await further research.

The empirical results on the outcome function tend to support the premises of the theory--for example, that outcome level rises with the

relative strength of the prosecutor--as well as to affirm the outcome function itself. The search for a point of discontinuity in the outcome function, anticipated because of the defendant's supposed specialization either in trial improvement or in consumption--yields positive results for all variants of the outcome function estimated. In spite of the roughness of the data, and of the somewhat precarious manipulations made in deriving the level of disposable resources for individuals, the results obtained tend to link higher income figures with lower plea sentence maximums. This may be reflective of a tie between defendant resources and plea outcomes or it may be ghosting a relationship which is more complex. It is possible that this variable serves as a proxy for the social and economic amenities of the various neighborhoods in which the defendants live and from which the aggregate Census data was drawn. In addition, this poses a number of questions about the interrelationships between income distribution and criminal activity.

Finally, in estimates of the resulting outcome function, the split of the Pareto optimal set is found consistently to be one favoring the prosecutor, in the ratio of seven to three. Although defendants receive sizable reductions in charge for pleading guilty, when we consider the amount that is actually "up for grabs," that is, the Pareto optimal set, the prosecutor faces relatively better in plea negotations. Comparing this finding with the fact that on average, potential sentences are reduced to about 38 percent of their original levels, would seriously raise the question of prosecutor "over-charging" in the prebargaining stages of the criminal justice process.

AN ECONOMIC MODEL OF PLEA BARGAINING IN THE CRIMINAL COURT SYSTEM

By

Judith Ann Lachman

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree

DOCTOR OF PHILOSOPHY

Department of Economics

1975

ACKNOWLEDGMENTS

I am deeply indebted to Professor William Haley, the chairman of my dissertation committee. To take on the job required courage; to pursue it required not only professional expertise, but also much patience and perseverance; and to execute it so well required action above and beyond the call of duty, the sharing of creativity, curiosity, and most insightful criticism. A debt of this sort cannot be adequately acknowledged, much less repaid; it can only be passed along to others--surely an awesome task--together with the encouragement and support so generously given to me.

The contributions of Professors Daniel Saks and Jan Kmenta to the development of both the dissertation and its author were vital ones too numerous to cite. These follow a long procession of gifts to me of their intellects, time, and enthusiasm spanning my graduate school career.

Professor Zoltan Ferency's resourcefulness and advice concerning the criminal justice institutions and law were invaluable both in the formulation of the theory and in the gathering and use of the data. He was always most willing to answer the often naive questions of a person uninitiated in the ways of criminal law.

My conversations and classes with Professor Warren Samuels constituted mental journies into many realms of economics; their legacy was a new set of intellectual tools, a renewed addiction to economics,

and a belief in the area of law and economics as a new territory to be explored.

My colleagues Stewart Hodges, Rich Cohn, and Franco Modigliani at M.I.T. shared with me their reactions and criticisms, and offered much encouragement along the way. I also benefitted greatly from comments offered at sessions of the Labor Seminar and the Economics and Law Seminar during my year at the Sloan School.

Data concerning Recorder's Court were made available to me through the cooperation of Presiding Judge John Murphy, Prosecutors William Cahalan, Dominick Carnovali, John McConnell, and Peter Bartholomew of Recorder's Court and the Wayne County Budget Office. Donald McIntyre and other personnel of the American Bar Foundation were most helpful in answering my questions and allowing me the use of their materials from the 1958 Study of Criminal Justice Systems and other items in the Foundation library. The assistance given by these persons was indispensable.

The art of persuasion, as applied to computers, was exercised and taught by my comrades-in-study Earl Knudson and Jan Palmer, who also lent their support and interest to the fruits of those labors.

The path to graduate study in economics has been a somewhat devious one. Much responsibility for my switch into economics belongs with the late Charles Ferguson, whose theory classes sparked my enthusiasm, and whose interest, persistence, and moral support made the change possible. I received much on-the-job training while working for the Bureau of Research at the State of Michigan; to my colleagues and "teachers" there--Drs. Vernon Smith, Gary Appel,

Bruce Stuart, Leland Hall and to Ronald Fine--I owe much in return. At that same time, classes and talks with Professor Milton Taylor provided a critical link with my academic endeavors. During my third year, the Center for Urban Affairs at MSU provided financial support as well as opportunities for teaching, research, and the general interchange of ideas.

In truth, the path to graduate study in economics--or at least, to somewhere--began much longer ago, and the persons who initiated it have been with me all along the way. To my parents--who have tolerated one crazy project after another, who have even been accomplices in some, who have shared in my joys and embraced me even in my failures--I bear a debt of much kindness and love.

Responsibility for remaining errors is all my own.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY OF SYMBOLS

SYMBOL	REPRESENTS	SECTION FIRST APPEARING
D	The defense	1
P	The prosecution	1
CC	Contract curve	2.1
S_1, S_2, S_3; S_α	Initial endowments of goods	2.1
K_1, K_2, K_3; K_α	Subsets of contract curve	2.1
I, II	Parties to exchange	2.1
r(S)	Relative strength of parties as function of initial endowment	2.1
a_j	Apples owned by party j	2.1
$(a_I(r_\alpha), -a_{II}(r_\alpha))$	Bargaining (Pareto optimal) interval corresponding to S_α	2.1
x	Object of distributive bargaining	2.1
yo	(Constant) level of all other goods	2.1
U(x,y)	Utility as a function of the amounts of x and y	2.1
U*(x)	Incremental utility function	2.1
Case V	The case P and D are both concerned with	2.2
Case X	A composite of all the cases other than V in the prosecutor's portfolio	2.2
v,x	Outcome of cases v,x, e.g., sentence measured in years	2.2
U=U(v,x)	Prosecutor's utility function	2.2

SYMBOL	REPRESENTS	SECTION
$B_{sf,v}$; $B_{sf,x}$	Costs per day of maintaining cases Vand X respectively	2.2
t_v; t_x	Number of days cases await disposition	2.2
T_v, T_x	Costs of executing a trial for case V; for case X	2.2
B	Total budget of prosecutor	2.2
B_f	Fixed costs of prosecutor's office	2.2
R	Disposable resources of P (assuming case V goes to trial)	2.2
F(v,x,R)=0	Trial opportunity set, stated as an implicit function	2.2
r'	Relative strength of D to P, as percieved by the prosecutor;	2.2
	larger values of r' reflect a stronger case for the defense	2.2
e	Initial endowment of evidence	2.2
d_i	Factors relating to the defendant, other than e, i=1,...n	2.2
$r'(e,d_1,d_2,\ldots,d_n)$	Relative strength of D to P stated as a function of e and the d_i	2.2
F'	The trial opportunity set, with r' assuming a higher value than before	2.2
$(v_o(r'),x_o(r'))$ and U_o	Solutions for utility maximization of P on trial opportunity set	2.2
t_d	Time until prosecution of case V is dropped by P	2.2
t_p	Time until case is resolved by a plea of guilty	2.2
G	Opportunity set when case V is resolved by a plea of guilty	2.2
R'=R'(R)	Disposable resources when case V is resolved by plea of guilty	2.2
G'	The guilty plea constraint with alternative values of one or more of its parameters	2.2

SYMBOL	REPRESENTS	SECTION
U'_o	Utility of P in equilibrium on the guilty plea	2.2
v_p, x_p	Values of v, x obtained when P maximizes his utility subject to G	2.2
r'_o	The initial level of r' on the trial constraint	2.2
r'_m	The level of r' exercised by D at which P would be indifferent between a plea of guilty and going to trial	2.2
v_m x_m	Values of y, x obtained via plea of guilty, for which P is indifferent between a plea of guilty and trial resolution	2.2
$I(v, x)$	The indifference curve (at utility level U_o) stated in implicit form	2.2
(r'_o, v_m)	Switch point (where v_m depends upon r'_o	2.2
$v_m = S_p(r')$	The prosecutor's switch function	2.2
α, β	For example given, parameters of P's utility function	2.2
k	Counterpart for case x of r'	2.2
$M_v, M_x; M$	Amounts of variable input M to production by P or a strong case	2.2
w	Price per unit of M	2.2
$\underline{M}$	Amount of good M equal to R/w (trial)	2.2
$\underline{M}'$	Amount of good M equal to R/w (guilty plea)	2.2
c	All other goods to which the defendant can allocate his resources, other than toward improvement of his case	2.3

SYMBOL	REPRESENTS	SECTION
V^*	The good V redefined to be a desirable good	2.3
v^*	Improvement in v relative to some benchmark	2.3
$\bar{v}$	A benchmark value of v, for example, the maximum value for D	2.3
R_T	The defendant's total resources, if going to trial	2.3
W_o	Wealth (or initial level of resources)	2.3
W_1	D's income not dependent upon jail status	2.3
Y	Income dependent upon jail status	2.3
J	Jail status: J=1 if D in jail, 0 otherwise	2.3
B_o	Defendant's fixed costs of trial	2.3
B_1	Defendant's time-dependent costs	2.3
$\underline{R}$	Defendant's disposable resources for trial	2.3
$f(v^*,c) = 0$	The defendant's trial opportunity set, stated in implicit form	2.3
r''	Relative strength of P to D as viewed by defendant	2.3
f'	The function f with alternative parameter values	2.3
c_o	c-intercept of f	2.3
v^*_o	v*-intercept of f	2.3
v^*_1	v*-intercept of f'	2.3

SYMBOL	REPRESENTS	SECTION
u_o^1	Utility of defendant, $(0,c_o)$ specializing in consumption (on trial constraint)	2.3
u_o^2	Utility of defendant, $(V_o^*,0)$ specializing in case-improvement (on trial constraint)	2.3
u_o	Maximum of $\{u_o^1, u_o^2\}$	2.3
r''_s	The value of r" such that $u_o = u_o^1 = u_o^2$; either type of specialization will yield same maximum utility to D on f	2.3
g	Defendant's plea bargain opportunity set	2.3
R'	D's disposable resources on guilty opportunity set	2.3
c'_o	c-intercept of g	2.3
u'_o	Maximum of $\{u'^1_o, u'^2_o\} = u'^1_o$	2.3
u'^1_o	Utility from c-specialization on g	2.3
u'^2_o	Utility from v*-specialization on f	2.3
$r''_{Promise}$	Level of r" exercised by P, which is promised to be less than r''_o	2.3
r''_m	Level of r" at which D is indifferent between modes of case resolution	2.3
r''_{disc}	Value of r" at which $u_o = u'_o$	2.3
S_{d1}	Portion of D's switch function (for $r \leq r''_{disc}$)	2.3
S_{d2}	Portion of D's switch function (for $r \; r''_{disc}$)	2.3
S_d	The entire switch function for D	2.3
r	The restatement of r' and r" as one variable	2.4

SYMBOL	REPRESENTS	SECTION
$S_p(r)$	New notation for the prosecutor's switch function	2.4
$U^*_o(v_m)$	Prosecutor's incremental utility function	2.4
$u^*_o(v_m)$	Defendant's incremental utility function	2.4
$\lambda; 1-\lambda$	Weights accorded the P's and D's switch functions, respectively	2.4
f_b, f_w	Outcome functions of a defendant, compared by race possibilities (in figure 25)	2.4
γ	$\gamma = 1$ for $r'' \leq r''_{disc}$ and $\gamma = 0$ otherwise	3.1
$h_p(r)$, $h_d(r)$	Functions which permit both switch functions to be stated in terms of a single variable r for relative strength of the parties	3.1
$\sum_{j=1}^{n} v_{oj}$	Sum of sentences on all pending charges (n is number of charges)	3.2
z_i	Characteristics of the defendant and of the offense (previously labeled "e" and "d_i")	3.2
$q(z_1, \ldots, z_d)$	A linear function of the z_i used in estimating the relative strength variable	3.2
r_p	Probability of conviction on original offense, as estimated by prosecutors	3.2
x_1	Term from P's switch function without λ coefficient	3.3
x_2	First term of D's switch function, without $(1-\lambda)$ coefficient	3.3
x_3	Second term (which includes dummy variable for regime) of D's	3.3

SYMBOL	REPRESENTS	SECTION
a_1, a_2, a_3	Coefficients of x_1, x_2, x_3, respectively	3.3
v_{mj}, a_{ij}, x_{3j}	The terms v_m, a_i, x_3 multiplied by the regression regime dummy variable Y_j	3.3
Υ	$r''_o - r''_{disc}$	3.3
k_j	Cuts in regression regime, j=1, ..., 10	3.3
Y_j	Dummy variables for parts of regression regime, j=1, ..., 10	3.3

1. Introduction

In contrast to popular belief and to the Constitutional right to trial by jury, the large majority of criminal cases in the United States are settled by prosecution and defense without trial.[1] Although clothed in controversy during occasional law school discussions and media tirades, the practice of plea bargaining has elicited little attention outside those circles and has continued to be the common mode of case resolution.

Legislation and judicial rulemaking have rarely addressed the issues of the controversy until quite recently,[2] and then primarily to give official approval to ongoing practice, rather than to actually create the institution in law. Plea bargaining has been--and it appears, will remain--a highly informal proceeding, allowing the parties great latitude both in structuring the negotiation process[3] and in determining an acceptable outcome.

In both its pre-bargaining posture and in its behavior during the course of the negotiations, each party pursues a special manifestation of its own self-interest. The defendant (D) seeks to minimize his expected loss resulting from the disposition of his case (time in prison and other forms of loss) and so he considers entering a plea of guilty in return for the dropping or reduction of some of the charges against him.[4] The prosecutor (P) attempts to maximize the total convictions among his cases,[5] together with his goals of disposing of all cases in the docket, and taking into consideration both the seriousness

of the charges and any other factors relevant to his priority judgement about them.[6] Although sacrificing the possibility of obtaining a more favorable outcome at trial, each party via pretrial settlement eliminates the risk of losing at trial, and frees his own time and additional resources for other purposes.

Both the prebargaining posture of the parties and the outcome of negotiation can be described as the results of economic processes in which individuals make rational choices among alternative courses of action on the bases of anticipated utility gains or losses. In the following pages a model is constructed which derives the prebargaining stance of each of the parties, and states the set of possible guilty plea settlements as a function of the relative strength of the prosecution and defense. The model is then tested using data on individual defendants in felony cases, with the resulting estimates serving both to affirm the model in a general sense, and to indicate the degree to which the parties are able to extract from one another the reservation price of a guilty plea.

This chapter serves to introduce some of the issues toward which the theory is directed and provides a sketch of the court system as depicted in the economics literature.

The second chapter is devoted to the development of a theory of plea bargaining: First the problem of indeterminate outcomes--the legacy of bilateral monopoly problems--is discussed, and a rationale for adoption of a split-the-difference theory is proposed for the bargaining problem in the context of the court system. In the following two sections, the choices of prosecutor and defendant are placed in the

traditional framework of utility theory, with each party allocating its resources between improvement of case outcome on the one hand, and all remaining goods on the other. For the defendant, the remaining goods are consumption goods, savings, and other uses of resources, and for the prosecutor the remaining goods are the outcomes on the rest of his caseload.

Each party chooses which of two alternative constraints to observe, where trial defines one of the constraints and a plea of guilty defines the other. However, if one party initially prefers the trial option, the second party can entice him to bargain by adjusting the "price" of a guilty plea--by agreeing not to exercise the full strength of his case in court--so as to make the first party indifferent between the two alternatives. Between the concessions necessary to make acceptable a plea of guilty are the set of pareto optimal outcomes. The determination of this set and the consequences of adoption of a split-the-difference theory of negotiation are discussed in the last section of the second chapter.

In the third chapter, the model is adapted for use of individual case data obtained from the Recorder's Court of Detroit. The first section outlines the additional assumptions necessary to ready the model for estimation. The following section describes the data and gives some information about the two parties to the case as well as introducing the construction of a variable for relative strength of the parties. The third section presents and discusses the results from estimation of the outcome function determined earlier.

Finally, the fourth chapter summarizes and interprets the

theory and empirical findings in terms of policy considerations and areas for further research.

Until quite recently, economists have devoted scant attention to issues of criminal justice, and within the new and growing literature on the economics of criminal justice, almost no research has been done on plea bargaining and related questions of court management.[7] One major exception to the latter generalization is William Landes's seminal article, "An Economic Analysis of the Courts," (1971) which is discussed below.

In beginning his paper, Landes promises first to identify the variables which relate to the choice made by P and D between a pretrial settlement (plea agreement) and trial. He then sets out to show that "the decision to settle or go to trial depends on the probability of conviction by trial, the severity of the crime, the availability and productivity of the prosecutor's and defendant's resources, trial versus settlement costs, and attitudes toward risk." (Landes, 1971:61) As a first attempt at expressing decisions of actors in the court system in terms of economic control variables, the article is a pathbreaker and provides a wealth of signposts for future research, some of which Landes has followed himself (1973). However, as a model of the trial-versus-guilty plea decision, his work is less convincing than if it relied upon a more realistic view of the legal institutions and the traditions surrounding them.

One major weakness in Landes's conception of these institutions is his perception of the relationship of the negotiated

settlement to the full trial, where the relationship involves not only the order of events in time, but also the role of the trial option in the the determination of the bargained outcome. In his model, Landes views the bargained outcome as entirely separate from the trial-versus-guilty plea decision, and he creates a model in which the latter decision is made first, independent of the result of the former process; that is, Landes's prosecutors and defendants decide first on whether to go to trial, and then later decide the conditions of settlement. This is analogous to a model of labor negotiations in which the parties first make an irrevocable decision on whether to have any strike associated with the new contract and then, if they happen to decide against striking, sit down to work out the terms of the contract.[8]

This conception of the trial decision has several crucial defects. First, as the comparison with labor negotiations suggested, Landes's model fails to recognize the intimate relationship of the conflict situation, trial, to the attainment of the negotiated agreement, the plea of guilty: the former indicates failure to achieve agreement. This last idea embraces two critical components, namely, that conflict signals a form of failure, and that conflict follows attempts at non-conflict settlement, rather than the other way around. In addition, Landes's conception fails to recognize the importance of the ability to choose the trial course of resolution--at least the ability to use the threat of going to trial--as a source of power in bargaining; for without the ability to reject a bargain proposal of the other party (in the world of an Edgeworth box, the ability simply not to exchange goods) neither party could induce the other to accept <u>any</u> proposal it might

make, and both could remain intransigent at the levels of their initial demands.[9] However, added to problems involving the logic of this assumption of the Landes model is the fact that the actors and the institutions which constitute their stage simply do not relate to one another in the way Landes describes. With the exception of a very few highly publicized cases, every case is viewed by a prosecutor as a potential plea bargain; and with the exception of a very few defendants, the decision about whether to go to trial results not from a simple decision to plead guilty because it is cheap, but from a comparison between the outcome anticipated from trial and the best negotiated outcome he could wrangle from negotiations. This sparring for position frequently continues until the very moment that a trial would otherwise begin; only after agreement has been consummated or finally rejected has the trial decision actually been made.

Finally, Landes makes the assumption that whenever a Pareto optimal bargain exists, i.e. whenever there exists at least one such outcome, then the parties will settle. This is to assume away the interesting part of the question,[10] namely, how do defendants decide to accept or reject a bargain and consequently avoid or participate in a trial? Because of the assumptions made and the underlying conception of the institutions involved, the Landes model is of limited assistance here in constructing a model of plea bargaining.

The view of the institutions and of the process of plea bargaining which are assumed in this dissertation are, briefly, as follows. Having gone through the preceding stages of the criminal justice machinery--arrest, investigation, and charging--a case enters the

prosecutorial stage, where the trial versus plea bargain choice is made by the parties. Although institutional practice varies from one court to another,[11] at some time while the case rests in the prosecutor's office the opportunity for pretrial bargaining arises. In the past this has frequently been carried out in hurried whispers before the commencement of trial. Now, in some courts such as in Detroit, the practice has been to institutionalize the bargaining session by arranging a pretrial conference as part of the normal processing of a case. In any event, it is fair to say that most any defendant in a large urban court has the opportunity to enter into negotiations if he wishes to consider pleading guilty.

Although some bargaining sessions are characterized by "hard-line" bargaining and the exchange of specific offers from one side or another, many are marked by a great deal of information-exchanging, which serves to bring about a consensus or the determination of a small interval within which most of the offers will reside. Although the trial option always looms in the background, discussion is centered on the special factors which mitigate the desirability of a harsh sentence (such as would be possible at trial on the original charge) and on the consequences of pleas to alternative charges.

Thus, once much of the preliminary information has been exchanged, the problem is essentially a search for the set of pareto optimal outcomes and then, via negotiation, a resolution within that interval. The theory of plea bargaining presented here rests upon this conception of the process by which a pretrial settlement is made, or rejected in favor of going to trial.

NOTES TO CHAPTER 1

[1]Estimates on the proportion of guilty pleas taken in criminal courts vary from 95 percent to 69 percent, according to McIntyre (1967:132). Newman (1966) reviews some of the studies producing those figures, and Morse and Beattie (1932) provide one example.

[2]For examples of these official stamps of approval, see People v. West (1970), Santobello v. New York (1971), and McCoy v. U.S. (1966).

[3]This is especially true for the prosecutor. As part of the executive branch in the criminal justice system, he can exercise discretion in the administration of guilty plea negotiations, subject to legal and political constraints. The concept and justification of prosecutorial discretion are discussed briefly in People v. West (1970) and in Remington (1969, 1972).

[4]The Newman article (1956) and book (1966) list and compare various modes of bargaining for pleas and the alternative mediums of exchange.

[5]These conditions placed on the utility function are justified by two considerations: Upon assuming his position, a prosecutor pledges "...not to convict, but to see that justice is done." Nevertheless, his conviction rate is frequently a primary topic of discussion during election campaigns, with a low one often cited by the opposition as evidence of a "soft" or "ineffective" job having been done by an incumbent.

A second reason for believing the prosecutor's utility to be as indicated is the reliance of the U.S. justice process upon the adversary system for the presentation of evidence and its interpretation. Such a system requires that a prosecutor scale a role barrier in order to introduce information which would tend to favor the defense at trial. The adversary model of the criminal justice system is discussed in more detail by Klonoski and Mendelsohn (1970).

[6]For example, the deterioration of evidence over time might impose time-varying preferences for the sequencing of cases through time.

[7]What has been done so far appears primarily in a new publication, the Journal of Legal Studies and in issues of the J.P.E. during the last four years.

[8]As another example, Landes's conception is similar to a belief that nations at war first end the conflict and then work out the terms of the treaty.

[9]This would be similar to the assumptions made (though not of specific institutions, nor of the relation of conflict to negotiated agreement) by John Cross (1969) in his model of the bargaining process.

[10]There should always exist a Pareto optimal outcome so long as one assumes the linear utility functions necessary for use of the von Neumann-Morganstern index (which Landes utilizes) because both parties save the money which would otherwise go toward the preparation for trial--and after-trial resources are the only consideration to Landes defendants and prosecutors. The outcome anticipated from trial, plus or minus quite a bit, would enclose an interval of Pareto optimal outcomes.

[11]Institutional variation is discussed in section 3.2.

2. A Model of Plea Bargaining: The Pre-bargaining Posture of Prosecution and Defense and the Outcome of Negotiations

2.1 The Problem of Indeterminacy

Within the body of microeconomic theory there arise a number of situations--most notably those involving bilateral monopoly or oligopoly--in which one is unable to prescribe a determinate solution to the outcome of an exchange process. Until the appearance of the early bargaining theories in the 1930's, the furthest an economist would take the analysis of such a relationship was to set boundaries on the outcome, based upon the outcome each party would have accepted in the competitive case, and close with a statement that although economic theory can suggest what the demands of the parties will be, the final outcome depends upon the strength of the bargaining parties, the latter being an issue outside the scope of economics.[1]

Such a statement is lacking in two important respects. First, as evidenced by the newly-burgeoning literature on bargaining contributed by economists, it is possible and indeed desirable that the discipline of economics countenance such issues. Second, the statement is misleading in its implication that, should other social sciences be able to furnish an economist with information on or quantification of the strength of the two parties, the economist would then be able to determine the outcome, i.e. that an evaluation of relative strength would be sufficient in the way of additional information to transform

an indeterminate situation into a determinate one.

These two difficulties became intertwined as most of the bargaining theorists set off in search of some measure of relative strength which could be defined wholly in economic terms.[2] This new variable could then be imposed upon the less-indeterminate-than-before "solution"--usually conceived of as an interval within which the outcome would lay --and yield an answer to the problem. The bargaining theorists essentially agreed with the earlier microeconomists, that the outcome of bargaining situations is determinate once a variable for relative strength is included; they merely disagreed as to which discipline should undertake the awesome task of measuring that variable. It is crucial to note that (1) with few exceptions, the bargaining theorists considered the model complete and determinate once a variable for relative strength had been included, i.e. whereas the microeconomists had considered the determination of the final outcome (within the previously determined bounds) to be entirely outside the scope of economics, the bargaining theorists considered it to be entirely within; and (2) the bargaining theorists did not perceive any problem of simultaneity in the determination of the bounds on the outcome and the relative strength of the parties. Both of these perceptions are evidenced in the development of the prototype theory: the existence of an interval of indeterminacy is presumed, with bounds already set by recourse to the competitive model; and then a measure of relative strength is introduced as rather a deus ex machina, resolving the dispute within the given interval.[3]

The natural result of such a conception of the bargaining

problem was a belief that if indeterminacy persisted even after the introduction of the strength variable, then one or another of the assumptions of the augmented model must not have been satisfied. Thus we see Hicks (1963), Foldes (1964), Cross (1969), and others investigating the effects of insufficient information on bargaining. Other economists explored the possibility that the negotiators are not optimizing their supposed objective functions, but are, perhaps, torn by differences of goals within their ranks--as Ashenfelter and Johnson (1969) assume for labor unions--or, as posited by deMenil (1971) that the two parties are not engaged in distributive bargaining pure and simple but rather experience a positive sort of interdependence such as the vested interest of a labor union in the profits of the company. In other words, the response to the persistence of indeterminacy of outcome in the augmented model was not to question whether the model was in fact determinate, but to search for reasons that the model might be inapplicable.

The pursuit of inapplicability was quite successful, and as a result, as noted by Hamermesh (1973), "Bargaining theory contains very few interesting propositions that can be tested empirically." This is due both to the restrictive assumptions eventually required (in an area meant to complement the restrictive assumptions of the competitive model!) for imposition of the bargaining theories and to the kinds of predictions the theories typically yielded, i.e. they frequently did not specify outcome in terms of the units of the good in question. Since some of the empirical work using bargaining theory will be discussed in a later chapter, we return now to the question of

indeterminacy within the framework of the model.

Although it is conceivable that one would encounter a bargaining situation in which the strength of the parties is independent of the other variables within the model,[4] we would expect this to be the exceptional occurrence. For the belief in the universality of that situation is equivalent to acceptance of the proposition that regardless of initial endowments of the bargain good or of input factors, regardless of the forms of the objective functions, and regardless of the levels of other variables in the theory, the relative strength of the parties remains unaffected at its initially-determined level; and conversely, that a given level of strength is in no particular way associated with the existence or size (however measured) of the bargaining zone.

In order to discuss this issue in a bit more specific a setting, consider the set of possible exchange relationships depicted by the Edgeworth box below:

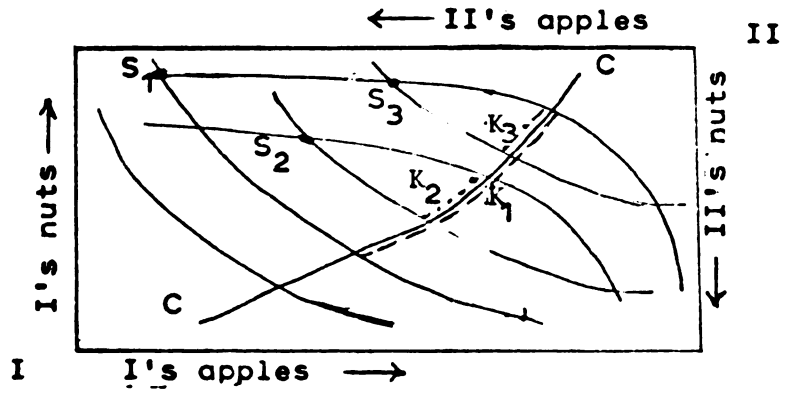


Figure 1. An exchange relationship: utility maps and outcome possibilities.

where K_1, K_2, and K_3 are the subsets of the contract curve CC corresponding to initial endowments S_1, S_2, and S_3, respectively, and the utility

functions of I and II are represented in the usual fashion. If we were to ask an unsuspecting subject to compare the relative strength of the parties at S_2 versus at S_3, we should not be surprised to receive the reply that I is relatively stronger at S_3 than he is at S_2. To give an exaggerated example of this, we might assume that at S_2 both parties have enough food to subsist, while at S_3, I considers additional food of either variety to be a luxury, and II is on the verge of starvation. If we define r(S) to be the relative strength of the parties, with r increasing with I's strength, we could restate our subject's conclusion as $r(S_3) > r(S_2)$.

If we take any line having a nonnegative slope (with respect to either I's or II's origins) we can order the endowment-strength variable r by the ordering of the corresponding values of one of the goods. For example, if we draw the line determined by S_2 and S_3, we have $\partial r/\partial a_I$ (where a represents apples) greater than zero, and in addition, for each S_α point on the line, we can determine a corresponding exchange interval K_α, and thus additionally create a correspondence of r-values with Pareto-optimal intervals:

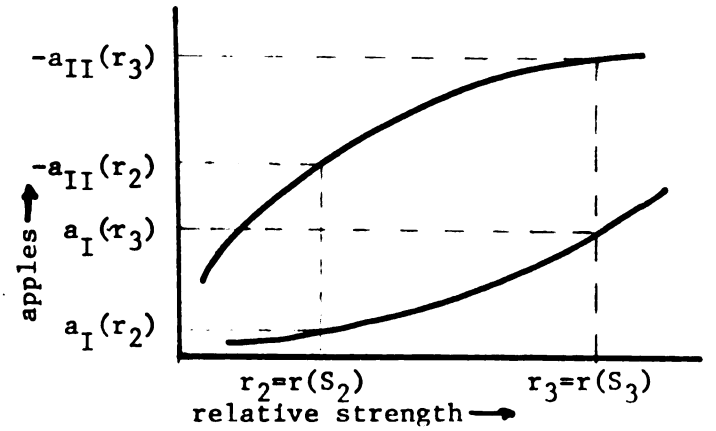


Figure 2. Reservation prices as functions of relative strength.

where ($a_I(r_\alpha), -a_{II}(r_\alpha)$) represents the bargaining interval corresponding to S_α. Note that when the set of goods to be divided consists of only one good, the line-ordering of the S_α has unambiguous meaning, and S_α is measured in units of the good under discussion.[5]

Such a sketch serves to underline some basic results of the Edgeworth-Box construct. First, it reiterates the fact that for a given level of relative strength--at least to the extent that it relates to endowments--there corresponds not one, but many possible bargain outcomes. Within the interval corresponding to a particular level of r, we would expect the final outcome to be a function of a number of variables, at least some of them relating to factors outside the scope of economics, such as the interpersonal skills and experience of the bargainers. We would consequently anticipate that any deterministic theory proposed by an economist would either take such considerations into account directly by including relevant variables or some general index to represent them; account for them indirectly by suggesting some approximation to outcomes which is sensitive to changes in the underlying behavioral variables; or content himself, with the aid of variables for which he does have information, to perform the valuable task of narrowing the interval of indeterminacy.[6]

Second, we note that a given bargain outcome can correspond to many levels of relative strength (although we generally expect a positive relationship,) and thus allow for the possibility of attaining an acceptable bargain even though the parties have incorrect or insufficient information, an occurrence which surely arises in the real world. For example, if I overestimates his strength at r_3 and also drives a poor

bargain compared to what he feels he could; and if II correctly estimates the relative strength to be r_2, they can both settle happily at outcome a_I in spite of their differing estimates. It is also curious--but consistent with the results of the Siegel and Fouraker (1960) bargaining experiments--that under some circumstances such as those of the illustration just given, increasing the information available to one or both of the parties may actually widen the bargaining interval and result in greater uncertainty of outcome, as well as possibly yielding a less favorable outcome to the party who was previously the more ignorant of the two!

As mentioned earlier, the deterministic bargaining theories address themselves to the distribution of an amount of a good within an already-specified Pareto-optimal interval. For the examples given above, the alternative to this distribution by means of negotiation was simply the failure to redistribute. In other bargaining situations, the "conflict" or nonnegotiated outcome may have associated with it various costs, outcomes, and levels of uncertainty relating to both costs and outcomes. Although the questions of relative strength and determination of bounds on outcomes remain the same, the problem of measuring these factors may become considerably more complicated. A number of bargaining theorists allow for this in shorthand fasion by the use of what we shall call an incremental utility function.

Very simply stated, the incremental utility function is a function of one variable--the object of the distributive bargaining, x--formed by stating the utility function as a function of x when x is

attained via bargaining (i.e. not by conflict) and subtracting the utility value associated with conflict, that is,

$$U^*(x) = U(x,y_o) - U(x_o,y_o)$$

where x is the outcome of negotiations, y_o represents the (constant) level of all other arguments of the utility function, $U(x_o,y_o)$ is the utility associated with conflict outcome x_o, and $U^*(x)$ denotes the incremental utility function. The latter (as well as the former) term may be difficult to summarize in a single dimension, but that is a necessary step. The incremental utility function is then a one-dimensional measure of the utility of x minus the constant utility value associated with trial.

Different theorists perform a variety of permutations on these incremental utility functions and the values they take on for particular bargain alternatives. For example, Nash (1950) predicts that, subject to a set of restrictive assumptions concerning the incremental utility functions, the outcome will be the distribution of x which maximizes the product of the incremental utility functions of the parties. Zeuthen (1930) hangs the fate of the parties of a pair of "risk willingness" expressions (later adapted by Pen, 1952) which Harsanyi has shown to yield the Nash outcome (1956). Cross (1969) has shown that the Cartter (1959) expressions for "bargaining attitude" are simply the inverses of Zeuthen's risk-willingness expressions, and hence yield the Nash outcome.

These theories share two common deficiencies: first, they fail to suggest why or how the parties could be expected ever to arrive at the predicted outcome point,[7] i.e. they give no hint as to the process or

dynamic involved in reaching consensus at that particular point; and second, they give the potential user of their theory no help in the difficult job of restructuring a specific bargaining problem in a manner amenable to application of the theory, particularly the task of developing the incremental utility functions.[8] Their solution to the bargaining problem, if not useless, is at best unsatisfying.

Several other theorists have suggested rather mechanistic resolutions to the bargaining problem. For example, Cross (1965, 1969) posits learning behavior on the part of the bargainers; but due to his assumption that both parties are completely intransigent in their required outcomes, his theory eventually rests on the contradictory[9] pair of axioms: (1) Although each party expects the other to accede completely to its demands in due course, it (2) then revises its expectations in response to moves by its opponent (but then repeats the naive assumption that its new demand will be completely fulfilled). Bental and Comay (1973) suggest an adaptive expectations model which is a bit more helpful than Cross's, but requires that the parties have a great deal of information at hand.

To summarize, each theory of bargain outcome discussed in the preceding pages bears one or more of the following deficiencies: (1) The theory neither relies upon nor is necessarily consonant with any concept of the dynamics of negotiation,[10] a criticism most readily leveled at the early (static) deterministic theorists; (2) The theory describes the process of negotiation quite well, but fails to predict outcome--just the opposite of the first deficiency--a problem especially common to the behavioral theorists and proponents of commitment stra-

tegies; (3) The theory is able to predict outcome and recognizes the stepwise nature of the bargaining process, but does not adequately account for the interaction between the parties, escaping that problem by effectively assuming away the two-party character of the bargaining. Since most of the empirical work to date has been tests of this third set of theories, the deficiency is particularly disappointing.

It appears that most of these theories have limited usefulness as descriptive models of the bargaining process, and consequently of plea bargaining. We are left with what might be called a basic indeterminacy: we would expect that in most bargaining situations, there is the possibility that two parties, both reasonably well-informed and rational, might disagree ex ante on the outcome of their negotiations. As pointed out by both Bishop (1964: 412) and Cross (1969: 33) (although in different contexts,) were that not true, we would find ourselves observing few instances where bargaining took place at all--for both parties would immediately accept the inevitable outcome and skip the bargaining.

An alternative approach to the bargaining problem which has been largely ignored[11] --perhaps because of its simplicity--is the straightforward maxim of splitting the difference between the reservation prices of the two parties, where the split is broadly construed to include any proportional split, not necessarily that of equal shares. A split-the-difference theory of bargained outcomes is consistent with both the behavioral theories of negotiation and with observed practice. Such a theory is also consonant with a concept of a negotiated outcome as the result of an interactive process of concession, based upon three

principal characteristics of bargaining: (1) Bargaining is executed by a sequence of alternating moves between the two parties; (2) the sequence of each party's positions is a strictly monotonic one, with the possible exception of the rejection of final offers; and (3) the amount of one concession depends upon the size of the preceding concession by the other party.

The first characteristic is simply a recognition that the eventual outcome is attained by a sequence of offers and counteroffers, each reflecting a new concession. This response pattern implies that the outcome is unlikely to be on the boundary of the pareto optimal zone, for that would require zero concession repeatedly by one party or the other.[12] It also implies that a commitment or "stay pat" strategy (Schelling, 1960; Saraydar, 1971b) employed by either party would make a conflict situation almost inevitable, so that instances we observe of negotiated settlements would tend to exclude situations in which a commitment strategy was used.

In addition to the bargaining process involving a number of moves by each party, "good faith" bargaining is generally considered as demanding some concession with each move. "Certainly, Party will consider who made the last move and whether that move was regarded by the other as an 'adequate' move...One side may not be happy with the amount of concession made by the other, but he is under considerable pressure to respond with some concession of his own." (Walton and McKersie 1965: 87)

Finally, the concession-response pattern requires that significant concession by one party be reciprocated by the other. That is, a certain

concession by one party implies a lower bound on the absolute value of the concession offered in the next move by the other party, meaning that there is likely to be some overall relationship between the size of I's concession and the size of II's; for example, it may be the pattern in negotiations between I and II for a two-unit concession from I to be matched by a three-unit concession from II. Summed over all moves, this response pattern would result in an outcome which is in roughly those same proportions.[13]

Although the characteristics cited above lend support to a split-the-difference theory as applied to a specified bargain interval, it is yet to be shown that a fixed proportional split should apply to any arbitrary bargain interval, given that it is valid for one. In fact, it is in doing so that the elusive nature of the relative strength concept is highlighted.

We have postulated a bargaining situation in which for a given pareto optimal interval the proportion of the difference accorded each party depends upon a set of behavioral and institutional factors. The interval to which the analysis was applied is associated with some value of r, the relative strength variable discussed earlier. The broader concept of relative strength has thus been introduced as a two-faceted concept, represented by strength as dependent upon initial position (the endowment strength variable) and strength as exhibited by the bargaining conventions and skill factor. These two manifestations of strength are not independent of one another, for clearly the value of the first determines the set of acceptable distributions of the good.

The bargain conventions and skill variable becomes more central to the determination of outcome once the endowment strength aspect has set the environment in which the skill factor enters.

Since the split of a given Pareto optimal interval depends upon certain traditions and upon the skills of the parties, the question is whether those factors influence the split of the bargainable interval differently for alternative choices of the intervals. Since neither bargaining conventions nor the interpersonal skills and experience of the bargainers is dependent upon the level of endowment strength,[14] the answer here supplied is no.

Thus the idea that the bargainable difference, whatever it is, is divided in some proportion which is uniform for alternative values of r would appear to be a reasonable basis for a theory of outcome. It is this theory that is adopted in the model of plea bargaining to follow.

We now face the task of defining the region of indeterminacy in the setting of the plea bargain-versus-trial choice, where the former is the result of negotiation, and the latter of the failure to reach a mutually acceptable bargain.[15] This is accomplished by developing the analogs of the incremental utility functions for zero increment, and allowing the relative strength of the parties to vary. For each, a "switch function" results, defining the minimally acceptable outcome necessary in order to induce the party to give or accept (as the case may be) a plea of guilty. The region bounded by the switch functions--for some ranges of relative strength, the empty set--defines the set of possible outcomes of successful negotiations. And within this region,

if one is willing to posit a rule or theory of outcome, the resulting plea bargain or trial option choice can be predicted.

NOTES TO 2.1

[1]For example, Charles Ferguson writes:

> Our general conclusion is that price and quantity is <u>indeterminate</u> in cases of bilateral monopoly...[T]he information the economist has is not sufficient to determine the precise market solution.

The additional information necessary to determine the outcome he considered to be "anterior to the realm of economic analysis." (1966:248) (underscoring in the original.)

[2]Pen expresses the need for an economic theory which is determinate, and does so in rather stark terms(1952:25):

> The term 'indeterminate' is used in this [mathematical] sense by most economists. It then means nothing more than that their system of equations, that is, their theory, is inadequate to offer an explanation of the magnitudes to be examined.
>
> A typical example of this is the 'theory of limits' of bilateral monopoly...It is superfluous to explain once more that the price is indeterminate within these limits; this merely amounts to saying that the investigation has been broken off before it came to a successful conclusion.

[3]Dunlop (1944; 119), for example, classifies the factors determining bargaining power into three categories:

(1) Tastes of workers and employees with respect to wages and manhours bought and sold, and institutional factors--such as property rights and wage-hour legislation--which influence demand and supply conditions;
(2) Market conditions, expecially the degree of competition in the labor market, the product market, and the market for complementary and competitive factors of production;
(3) Pure bargaining power: the ability to get favorable bargains apart from market conditions.

[4]For example, suppose that a bargaining interval is determined entirely by technical factors, that disinterested negotiators are hired and paid on identical incentive schemes for each unit of the good above or below the appropriate bound which is attained at bargaining, or for obtaining a conflict outcome. Further, we must either require that each level of relative strength be associated with bargaining intervals of fixed length, or specify attainment in terms which adjust to the length of the bargaining interval.

[5]Note that otherwise it is necessary to also order S_1 relative to S_2.

[6]Cross scoffs at this last role, since in his view, the issue is simply one of lack of information (1969: 25, 28)

> This assumption of perfect knowledge is not really a simplification of the bargaining process; indeed, we take it to be a recitation of one of its most fundamental properties. It is precisely the lack of knowledge that permits two individuals to form differing opinions as to how a situation which requires agreement will turn out...An assumption of perfect knowledge will reduce the whole process to a set of mechanical rules from which we cannot possibly gain any insight into the dynamic mechanism of concession, or into the process whereby the parties formulate and change their expectations. (underscoring in original)

[7]For example, the Nash solution is not a solution to the bargaining problem, but merely narrows the set of possible outcomes, and does that without reference to any dynamic or process which would cause the parties ever to arrive on that function--or even to have a motivation to do so. As Comay et al have pointed out (1974), the Nash solution has come to be regarded more as a normative solution than as a positive one (thus rescuing it from the responsibility of being empirically verifiable). However, even this kindlier interpretation of his theory leaves open the question of why that solution would be viewed as preferable by the "exchange community" of the two parties, by some ephemeral outside world (i.e. socially desirable to a larger community, if not optimal in the incomplete model of two-party negotiation), or to either of the parties. Surely each party is maximizing his own utility, and within the Pareto optimal region, would strive to gain an outcome on the other party's switch function (in the terminology of labor negotiations or oligopoly, at the reservation price of the other party) and only in the exceptional situation would this coincide with the Nash function. As a community of two, P and D--or any other two bargainers, for that matter--would have little reason to maximize the product of their utility gains from bargaining if both could be happier with an alternative distribution. Finally, we have no hint as to why some sort of outside world might find that solution to be socially desirable.

[8]The unsuspecting user of the theory may well find himself faced with one of the following problems: (1) The incremental utility functions only lend themselves to statement as implicit functions, thus rendering the theory's required manipulations extremely cumbersome, if not impossible; or (2) Solutions (distributions of the good which is the objective of the bargaining) that satisfy the conditions of the Nash outcome may not exist, or if they do exist, may not be real numbers. Further, the fact that for certain pairs of utility functions the outcome must always be on the boundary of the pareto optimal interval revives one's reservations about the role of the variable for relative strength in the determination of that interval, and of the outcome within the interval.

[9]There is some disagreement concerning whether the Cross behavioral assumptions are contradictory or merely paradoxical. The controversy is exemplified in Coddington's review (1970) of the Cross book.

For whatever relief it affords, it is here noted that the assumptions cannot be contradictory in the first period, at least.

[10]This is analogous to offering a point as a market equilibrium without clear definitions of the demand and supply functions, or without a cobweb theorem offering a reasonable explanation for the equilibrium point being an attainable one.

[11]The one notable exception to this is a study done by Daniel Hamermesh (1973) in which the author, while attempting to test the Nash theory, instead tests for a fifty-fifty split of the difference between labor's and management's first offers. Since he is testing for the applicability of the Nash theory (which requires that if the two parties have identical utility functions, then they must divide the pareto optimal interval in half) and finds it to be probably inapplicable, he does not test the Nash theory itself. In Hamermesh's opinion, the Nash theory is a split-the-difference theory--in the admittedly unlikely circumstance that it should be applicable to a bargaining situation--and so he believes that he has tested the Nash theory as well as a split-the-difference theory. I briefly discussed that interpretation with Robert Bishop, who disagrees with Hamermesh, stating that the Nash theory is not a split-the-difference one. To support Bishop's view: the Nash theory rests upon four stated axioms and (depending upon one's approach to assumptions made en route to a theorem) two or more assumptions less prominently set forth. The failure of any particular bargain situation to satisfy an axiom implies that the Nash theorem on outcomes cannot be applied; this is because the antecedent conditions have not been satisfied, and follows from simple logic. Hence it is here contended that the result of the Hamermesh test--even were other methodological problems absent--is not to test the Nash theory, but rather to show why that theory cannot be tested on his data with his assumed utility functions.

[12]Walton and McKersie (1965: 92) point out that "The most difficult aspect of making a concession at any point during negotiations is to be sure that it is coupled to a concession from Opponent."

[13]In this example from Ann Douglas's study (1962: 278) a concession by the union has effectively telescoped several moves, thus requiring similar action from management. A mediator cautions management to resist following their initial plan of small-step concessions, in order to respond to the union in kind:

> Now look. That union has got right into the realistic area. If you come up with a cent or 2¢ [sic] you are going to damage these negotiations beyond repair. Now you know you have to settle this and you know that you have a--a limit.

> The question if--I know what you're trying to do. You're trying to get to that limit without jeopardizing your bargaining position, and I'm here to help you do that, but don't make it harder on me. The union has made such a substantial move, and now you know they're in the realistic area. If you come out with a penny, you're gonna insult them. You're goin' to make them feel that you're not bargaining in good faith. If I were you I would throw out at least 80% of what you're going to give them--and I didn't know what you were--what the company was going to give them.

[14]However, for alternative values of the endowment aspect of the strength variable--i.e. violating the underlying ceteris paribus assumption--there may be differences. For example, the particular negotiators on both sides of General Electric contract bargaining may be considerably different than those who would be hired if G.E.'s strategy were to be radically different from the bulourism which has characterized it in the past.

[15]This is in contrast to Landes, who conceives of a bargain as the failure to go to trial. Fortunately, he is alone in his perception of bargaining as the failure to have a conflict. Unfortunately, with a few exceptions--Hicks, Cross, and the nondeterministic theorists such as Walton and McKersie--few perceived of conflict as signifying failure in negotiations.

2.2 The Prosecution

The prosecutor, in addition to his role as one of the parties in the trial arena, fulfills an administrative function in the judicial process. Having limited resources, he is confronted with more cases than he can investigate thoroughly and pursue fully through the complete criminal justice process. Consequently, he must make a choice as to the allocation of his resources among the various cases,[1] so as to obtain the best outcome overall, or effectively make such a decision by default.[2]

His utility is then a function of the outcomes of the set of cases before him:

$$U = U\ (v,x)$$

where v denotes the outcome of Case V, which we single out for detailed consideration, and x represents a composite of the outcomes of all the other cases under the assumption that the prosecutor (P) allocates his resources among those remaining cases in an optimal fashion.[3] (Here, outcome may be measured in a number of ways. For example, it might be the actual sentence given, measured in years; or it might be an index of various aspects of the outcome, such as a weighted average of a fine and a prison sentence.) The prosecutor (P) then seeks to maximize his utility subject to the constraint imposed by his budget and his ability to use the resources at his disposal in a productive fashion.[4]

The prosecutor's utility is then a function of two variables, v and x, and we make the usual assumptions about the shape of his utility function with respect to these variables, i.e.

$$\partial U/\partial v > 0, \qquad \partial^2 U/\partial v^2 < 0,$$

$$\partial U/\partial x > 0, \quad \text{and} \quad \partial^2 U/\partial x^2 < 0.$$

Given his utility function, the prosecutor confronts a constraint on his total resources: his budget for the relevant time period (and, although alternative values for the budget allotment are considered below, it is assumed that whatever its value, B, it is fixed for the time period under consideration.) Within his budget, P finds himself subject to certain fixed[5] costs, for example those associated with owning and maintaining the physical facilities, certain "housekeeping" personnel (i.e. those not directly involved in the handling of cases) and other commitments which are his legacy from the preceding period.

The remainder of his budget is devoted to costs which are variable in the sense that they can be attributed to the processing and disposition of particular cases, for example, the time of attorneys and other supporting staff. However, use of the term "variable" is a bit misleading here for the following reasons. The total amount of these resources is fixed (within the time period we are considering, for example, the budget year,) since they are defined to be the difference between total resources--which are fixed--and costs which are within the traditional definition of fixed operating costs. Hence, "variability" might be more accurately labeled "attributability".

Within the total of "attributable" resources are three sets of costs.[6] Considering the costs involved in the processing and disposition of a single case (at least to the extent that P contributes to determination of the disposition) we first note the costs associated with the processing of a case, independent of outcome. Specifically,

if $B_{sf,v}$ and $B_{sf,x}$ are the costs per day of maintaining cases V and X respectively, and t_v and t_x are the number of days to which the former costs apply, the attributable resources are reduced by the amount

$$t_v B_{sf,v} + t_x B_{sf,x}$$

where the case-maintenance costs are what will be called semi-fixed costs--costs which do not vary with output, but do vary with time.

The second subset of attributable resources is the set of costs associated with the execution of a trial;[7] these are labeled T_v and T_x, and their sum further reduces the available resources.

The remaining resources are ones that are truly disposable resources: those which P can choose to allocate to the improvement of one case or another. Denoting the total budget by B, the fixed costs of the prosecutor's office by B_f, and disposable resources by R, we have that

$$R = B - B_f - t_v B_{sf,v} - t_x B_{sf,x} - T_v - T_x.$$

Assuming that all cases come to trial,[8] it is only the disposable resources R, which the prosecutor is free to spend in pursuit of the maximization of his utility, the other portions of his budget having been already committed to the various costs mentioned above.

Before discussing the trial opportunity set which relates to R, let us consider some ways in which the level of R is affected by alternative values of its component variables. First, and most obvious, R is increased by increases in the total budget, B, or decreases in the fixed costs of running the institution, B_f. Second, cuts in pretrial delay for v or x, or reduction of the maintenance costs for the trial

docket would result in a higher R. Finally, reduction in the costs of trial(s) would make available a larger amount of disposable resources. Although scant reference will be made in future pages to the determination of R, it should be noted that alternative values for the underlying variables can be expected to effect shifts in R and hence yield alternative trial opportunity sets.[9]

Within the limitation of his disposable resource, R, the prosecutor can choose to allocate his resources between X and V, and we would expect there to be some tradeoff between the two. That is, if $F(v,x,R)=0$ is the trial opportunity set, stated in implicit form, we anticipate that

$$-\frac{\partial F/\partial v}{\partial F/\partial x} < 0$$

(applying the implicit function theorem.) Should F be a linear function, as in the example at the end of this section, the left-hand expression in the inequality would be simply the slope of the constraint.

However, we cannot expect F to be linear in the general case, for the following reasons. The units of v and x obtained by commitment of R resources are not goods purchased by P at a constant unit price, but rather, each case outcome is the result of a production process involving inputs such as the initial endowment of evidence, additional sleuthing by or on behalf of the prosecution, contributions to the strength of the case made by expert witnesses and technical resource persons, the skilled efforts of the prosecuting attorney, etcetera.

If we assign to this production-of-a-strong-case function the same attributes we typically assign to other production functions--positive and diminishing marginal productivity with respect to each fac-

tor--we obtain cost functions which bear the usual properties. If we then further require that the total amount of resources devoted to v and x be equal to R, we can express the set of attainable pairs (v,x) with x as a function of v. It is this function--actually, a production-possibility frontier (ppf)--which was called F(v,x,R) above.

For our purposes, F has two parameters which are of special importance. The first of these, R, has already been discussed. The second, which we label r', is a measure of the relative strength of the defense (D) to the prosecution, with larger values of r' reflecting a stronger defense case than do lower values. (In the special case of a linear opportunity set, r' is proportional to the absolute value of the slope of the line.)

Relative strength, as measured by r', is actually many factors in disguise, since prosecutors make their estimates of relative strength based upon a number of considerations: r' represents the subjective evaluation by P of the strength of the defendant's case relative to that of his own.[10] Included in this estimate is, at least initially, an educated guess as to what the other party is likely to do in the way of case development, as well as a hopefully more accurate guess about what he himself is likely to do in his preparation and presentation of the case at trial.[11]

Subjective as such an evaluation may appear, it is possible and, according to attorneys from both sides, the usual practice,[12] to make exactly such appraisals based upon certain objective criteria as: the previous criminal record of the defendant; various demographic characteristics of the defendant, such as race, age, sex, relationship to the

labor force, jail status; and other factors. In addition, they consider the endowment of evidence--and in particular, the initial endowment of evidence received by the prosecutor from the police, such as positive identification of the defendant by witnesses, legally obtained confessions, etc. These are items of information which could be brought out at trial, especially if the defendant should take the stand and testify in his own behalf, and are factors which tend generally to favor one side or another.[13] Letting e represent the initial endowment of evidence,[14] and the d_1 the other factors relating to the defendant,

$$r' = r'(e, d_1, d_2, \ldots, d_h).$$

Clearly, alternative states of these variables can be expected to affect the level of r'.[15]

The effect of higher values of r' is to associate with each value of v a lower value of x.[16] This is illustrated graphically below by F':

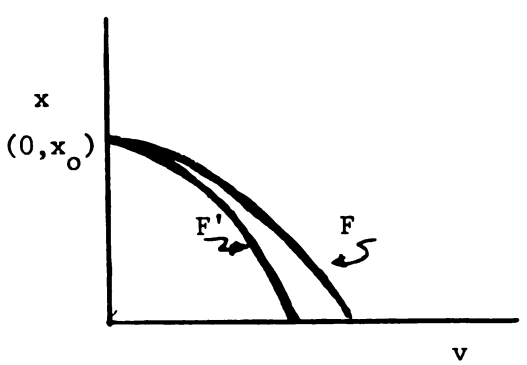


Figure 3. The effect of alternative values of r', the relative strength of D to P.

It is now possible to describe the prosecutor's allocation decision in terms of the familiar utility maximization, subject to a constraint which is possibly nonlinear.[17] Let $F(v,x,r')=0$ be the opportunity set, stated implicitly, on which we consider alternative (but predetermined) values of r'. If $U(v,x)$ is the prosecutor's utility function, we seek to maximize U subject to F. Since we have assumed that the opportunity set is (nonstrictly) concave with respect to the origin, and that the utility function is of traditional form, the outcome of P's constrained maximization problem yields a unique solution, $(v_o(r'), x_o(r'))$ for case outcomes, and consequent level of utility, U_o.

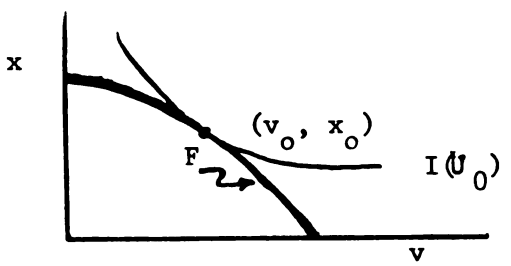


Figure 4. Prosecutor equilibrium, taking case V to trial.

The effects of changes in r' upon the prosecutor's allocation of resources and upon expected case outcomes can be traced through their effects on F, to determine alternative values of v, x, and U. Decreases in r', however incited by one or more of its contributing variables, (as well as increases in R) should result in a higher value of U, and of v, x, or both.[18] For example, the existence of previous arrests of D, other things being equal, should thus result in a lower value of r' and a consequently higher value of v and U than would the absence of such a

record. Similarly, information that the defendant is nonwhite, or of other factors favorable to the prosecution would cause the value of r' to be lower than otherwise.[19] Increases in e are unfavorable to the defense, and hence lower r' and raise v.[20]

In analogous fashion, reduction in pretrial delay, or of the costs of trial (for example, recent experiments on the use of juries of six members) would result in a lower value of R, shift F outward, and would lead P to adjust his allocation and attain a higher expected value of v and U,[21] as illustrated by F' below:

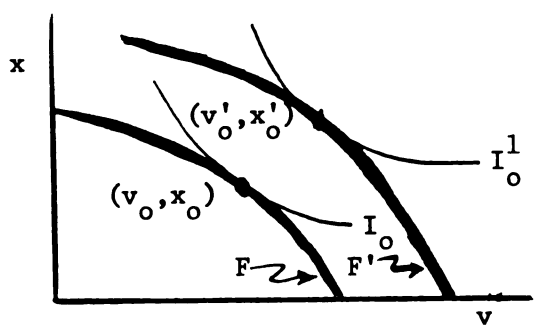


Figure 5. Effect of change in disposable resources R.

Thus it is clear that the cost effects of pretrial delay are unfavorable to the prosecutor.

In addition to the cost effects, however, the passing of time may have evidence effects as well. Having been endowed with a cache of evidence or information by the police, P faces the possibility of partial loss of that evidence as a function of the time interval between the offense or gathering of evidence, and the time of its utilization at trial. It is generally agreed that evidence in favor of either

party diminishes with the passing of time, although it is conceivable that it could increase, independent of any evidence-gathering effort by a party.[22]

The relevant question, then, is not whether usable evidence declines over time, but rather what the impact of this decline is upon the parties relative to one another; and the net effect is generally construed to favor the defense.[23] The evidence effect of delay is represented in the construct thus far developed as a decrease in e and a consequent increase in r', since r' is a function of e.

The sum of the two time (delay) effects is determined by the reduction in R due to the increased case-maintenance costs, and the increase in r' due to the decreased availability of earlier existing evidence.[24] The result of these two effects is to assign to each value of x a lower value of v than would have been assigned with a shorter delay period, represented graphically by a downward shift and/or clockwise rotation of the opportunity set F:

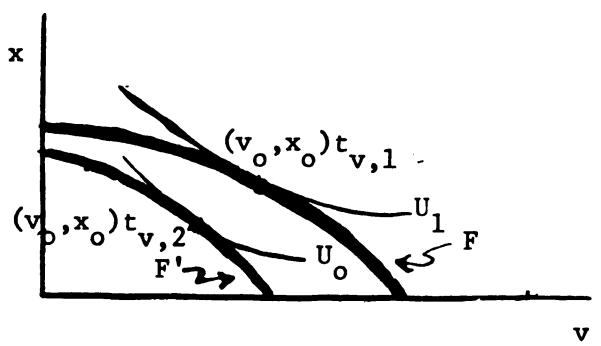


Figure 6. Time delay effects.

Consequently, the prosecutor obtains lower values of v and U, (with the possible exception--rather unlikely--of V being an inferior good in the relevant range of R.)

So far in this section, a construct has been proposed to explain allocative decision making by a prosecutor, who with limited resources attempts to obtain the best possible outcomes v and x at trial. Differing factors relating to each case--the initial endowment of evidence, expected time until trial, and characteristics of the defendant and of the offense, as well as P's ability to develop a strong case based on these factors--determine the opportunity set which is open to him for a given level of disposable resource. However, this description of a prosecutor's choices is unrealistic for two interrelated reasons: First, if a prosecutor were to bring every case to trial, he would be unlikely to have any disposable resources left for making the decisions we have described; in fact, he would probably run out of resources before disposing of even a quarter of the caseload.[25] The natural result of such a shortage of funds is to drop some cases in order to do a better job with the others. The power to make such selection is within the scope of the prosecutor's discretion and this discretion is exercised at a number of levels of the criminal justice process prior to the judicial stage.[26] If P were to drop case V, for example, at time $t_d \leq t_v$, the scheduled time for trial, the effect would be to increase P's resources by

$$B_{sf,v}\,(t_v - t_d) + T_v \quad .$$

This new level of resources could be allocated among the remaining

cases, while V would be assigned the value v = 0 (unless the indexing of v were so constructed as to include the psychic costs of waiting for an outcome, in which case v might be greater than zero even for cases ending in a prosecutor's drop, dismissal, or acquittal.)

A second option for disposing of cases other than the full-trial alternative originally introduced, is for some of P's cases to be resolved by a defendant's plea of guilty, thus avoiding the cost of trial and the more extensive preparation which trial would require of P.[27] In addition, in some courts, separate machinery is set up to negotiate and process pleas of guilty, thus considerably reducing the time from the offense to disposition of the case. In the event of a plea of guilty at time t_p R would increase by

$$B_{sf,v} \cdot (t_v - t_p) + T_v$$

where we assume that the trial time necessary for the defendant to make his plea is sufficiently small to save P virtually all of T_v. In the event that D pleads guilty, $v = v_p$. If we further allow v_p to range as low as zero, the first case, that of a zero non-trial outcome, can be subsumed under the second, and viewed as a limiting case. Were it not for these two avenues of flexibility, the prosecutor would have scant ability to pursue any of his cases, even high priority ones, beyond the outcome attainable at zero level of disposable resources.[28]

What are some of the characteristics of a guilty plea, from the point of view of the prosecutor, and how ought these to be interpreted within a model such as the one unfolding here? The typical resolution of a case by plea of guilty involves the defendant's entering of the

plea in return for either a promise of leniency in sentencing (where the units of v are measured as, say, the maximum penalty possible,) or a reduction or dismissal of some of the charges against him. Which of these alternatives occurs depends upon the particular court in which the case would be tried--i.e. restrictions imposed by applicable state, local, or federal laws, internal rules, the institutionalization of certain traditions (such as the belief that a reasonable bargain is the reduction from charge "Z" to charge "Attempted Z")--and within those restrictions, upon the discretion of the prosecutor.[29]

The view of many prosecutors that accepting pleas of guilty constitutes "compromising with the defendant" indicates a belief that the bargained outcome is in general less than the expected outcome of trial, although perhaps not uniformly so. Nevertheless, in cases resolved by guilty pleas, the utility to the prosecutor of that outcome must exceed that of the trial alternative, or he would have insisted upon trial. Considering the outcome opportunity set to be a function different from F--labeled G, and referred to below as the plea opportunity set--we can make some inferences about this new function and its relation to the trial opportunity set. First, because of the saving in trial costs and possibly from reduction in pretrial delay, the level of plea resources $R' = R'(R)$ exceeds R, and hence the x-intercept of G exceeds its counterpart on F. (The relation between the F and G values for the v-intercept will depend upon r' and other factors mentioned below, and hence is not determined at this point.) Second, G must dominate F (i.e. lie to the northeast of F) for at least some values of v--or else G

would never be preferable to F, and no rational prosecutor would find occasion to accept a guilty plea. Coupled with our first inference, that the x-intercept of G is greater than that for F, the second would suggest that G dominates F for low values of v, if not for higher ones:

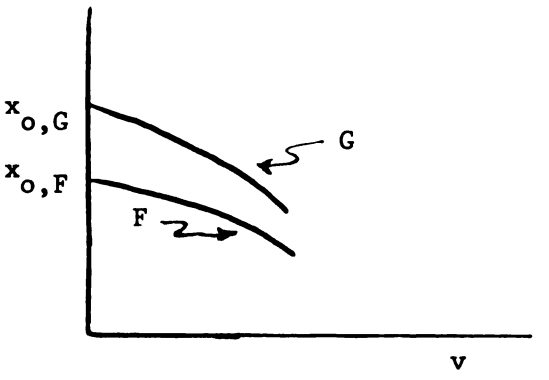


Figure 7. Positions of F and G for low values of v.

Finally, the fact that prosecutors believe most plea bargains to result in a lower value of v than would trial, implies that for some values of v, F must dominate G;[30] for (resorting to proof by contradiction) were that not true, i.e. if G always dominated F, a plea bargain would result in the typical reduction of v only in the event that P considered V to be an inferior good.[31] Thus for some values of v--to be compatible with the first two inferences, for some large values of v--F dominates G:

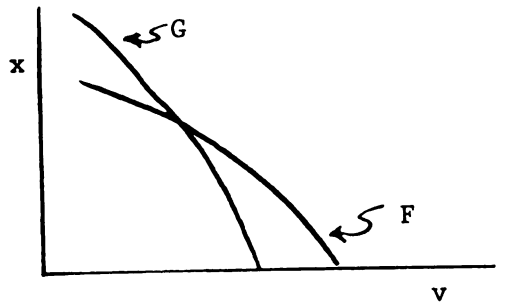


Figure 8. Positions of F and G: the v-intercepts.

This last statement, and particularly its graphical interpretation of the v-intercepts, is not surprising: it says that the best outcome obtainable from trial is at least as great as the best obtainable from a plea of guilty.

The set of factors which determine G is the same as that which determine F, except that the difference between t_v and t_p may enter into the determination of F for some defendants in some courts (in particular, if the court maintains separate pre-bargaining and pre-trial queues.) Given the values of the relevant parameters, it is possible to state F and G, and one may even say that G is the plea constraint which corresponds to F, since a change in any of the values of the contributing parameters would yield a new pair of functions, F' and G'.

As he did earlier for F, P determines the highest utility level he can reach while observing G as a constraint, i.e. he determines values of x and v such that

$$\frac{\partial U/\partial v}{\partial U/\partial x} = \frac{\partial G/\partial v}{\partial G/\partial x}$$

In addition to obtaining the equilibrium values for v and x, the prosecutor can associate with the plea option a utility value U_o', which may or may not compare favorably with U_o, the utility value he associates with trial. Clearly, if $U_o' \geq U_o$, P will choose the plea of guilty as his preferred method for disposing of the case. An example is sketched below.

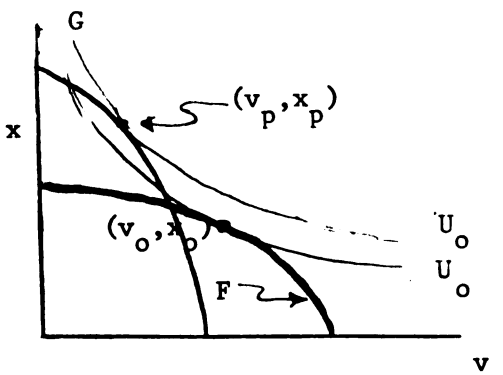


Figure 9. Prosecutor equilibrium on trial and guilty plea constraints.

If $U_o' < U_o$, the trial option would be the more attractive one for P, if the only alternative to F is believed to be G. However, G may not constitute the only alternative to the trial constraint, once we allow the possibility of the defendant's not bringing to bear the full strength of his case, as he would otherwise be expected to do at trial. Instead, in pleading guilty the defendant typically presents very little evidence of his innocence. Recall that both F and G were derived under the assumption that the relative strength of the two parties is r'. Now suppose that, although the value is r', the defendant were to agree as part of a plea bargain, to not exercise the full strength of his case; that is, he promises to use only r' which is less

than r'_o. Applying the earlier discussion of the effects on the opportunity set resulting from alternative values of the parameters, such a pledge by D would result in a counterclockwise rotation of G, and possibly result in making a plea of guilty preferable to trial.

The above interpretation of the kind of compromise involved in a plea bargain is quite straightforward. Typically, when the two parties go to trial, they "throw everything they've got" at the other party and consequently, r' can be properly construed to be the appropriate level of strength used in the derivation of F. However, the evidence presented at the time of a guilty plea is not at all the volume from either party which would be mustered for trial upon a plea of not guilty--else there would be staged a full trial, in effect. In fact, each party will present very little, usually just enough to suggest that the charge to which the defendant is pleading is an appropriate one for the circumstances of the case.[32] By not presenting his strongest evidence at the time of pleading, the defendant who pleads guilty has agreed to exercise a low value of r' (as viewed by P) in spite of the fact that the true value (the maximum, i.e. r'_o) may be higher.[33]

The extent to which D would have to reduce r' in order to induce P to consider a plea of guilty--that is, the magic level of r', r'_m, necessary in order to make P indifferent between observing F based upon r'_o and observing G based upon r'_m--can be determined within the present construct. Since F remains fixed regardless of the value of r'_m (in fact, r'_m depends upon r'_o) the utility of trial outcome, U_o, remains constant, and provides a benchmark for evaluating the relative attractiveness of plea outcomes resulting from the values of r'_p which are under consideration in the search for r'_m. U_o thus determines a

function $I_{r_o'}(v,x) = 0$, the equation of the indifference curve at that utility level. To determine the appropriate value of r_m' we restate G to include r' explicitly (although r_m' is not a variable, but only one possible value of the parameter r'.) With the plea opportunity set now written as $G(v,x,r') = 0$, we seek v_m, x_m, and r_m' satisfying the following conditions:

$$\text{(i)} \quad I_{r_o'}(v_m,x_m) = 0 \ ,$$

$$\text{(ii)} \quad G\ (v_m,x_m,r_m') = 0 \ , \text{ and}$$

$$\text{(iii)} \quad \frac{\frac{\partial I}{\partial v}(v_m,x_m)}{\frac{\partial I}{\partial x}(v_m,x_m)} = \frac{\frac{\partial G}{\partial v}(v_m,x_m,r_m')}{\frac{\partial G}{\partial x}(v_m,x_m,r_m')}$$

That is, we seek the value of r_m' and coordinates (v_m,x_m) for which the adjusted constraint G is tangent to the indifference curve $I_{r_o'}$. The solution is graphed below:

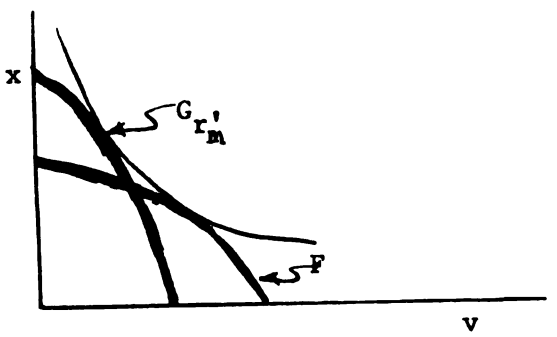


Figure 10. Prosecutor indifferent between the modes of case resolution.

To briefly recapitulate the geneology of r'_o and its descendants, we note that r'_o is a given value of the variable representing relative strength in case V as viewed by the prosecutor. This value, along with a corresponding one for X, the composite of P's other cases, are parameters in functions which describe the production-of-a-strong-case-for-trial. Given a fixed amount of disposable resources, R, and a suitable translation from units of the various inputs for strong-case production, the output X can be stated in terms of the output V; and this function, a production-possibility frontier, is the function $F_{r'_o}(v,x) = 0$. A similar function, $G_{r'_o}(v,x) = 0$ describes the opportunity set available if the case is resolved by a plea of guilty, rather than by a full trial; although R enters G as a parameter, it does so in the guise of R'(R), the disposable resources level made possible by savings from bypassing trial and perhaps a shorter pre-disposition queue.

When P performs his utility maximization routine subject, alternatively, to F and to G, he chooses to observe the opportunity set which affords him the higher level of utility. However, there exists a level of r', namely r'_m which, when substituted for r' in G, causes adjustment of G such that P is indifferent between disposing of the case via F or via G. At this point, one can consider a number of juxtapositions, one with another, of the many special values derived by the process just outlined. For the initial value r'_o yielded v_o, x_o, and U_o when subjected to F; and U_o in alliance with G led us to values r'_m, v_m, and x_m. Although we could choose to represent this set of related values by an ordered n-tuple $(r'_o, v_o, x_o, U_o, r_m, v_m, x_m)$, it is the relation of r'_o with v_m in which we have especial interest. We refer to the ordered pair (r'_o, v_m) as a "switch point" for it expresses the minimum level of

v necessary (when obtained by a plea of guilty) to pursuade P to accept a plea of guilty rather than to pursue the route to trial.

Two other pairings of entries from the seven-tuple are of passing interest here. One is the difference, $(r'_o - r'_m)$ which describes the compromise necessary from D in order to induce P to consider acceptance of a guilty plea as a viable alternative. Of course, D may eventually concede a great deal more, or none at all, as the bargaining progresses; $(r'_o - r'_m)$ only specifies the initial compromise necessary to convince P that D is quite serious about a plea.[34] Second, we take note of the difference between v_o and v_m, which measures the difference in outcome units which P experiences by switching from F to $G_{r'_m}$. Whereas the difference $(r'_o - r'_m)$ represents sacrifices by D, such as the promise to not exercise his rights to trial (including judgement by panel of peers, the right to confront witnesses, etc.,) a weakened position if he should decide to appeal the verdict or sentence, and others, here $(v_o - v_m)$ represents the associated outcome loss to P.

Having shown that for a given level of relative strength (as perceived by P) we can derive the corresponding plea bargain constraint and level of v_m which would make the prosecutor indifferent between trial and settlement, it remains to be shown how these items--and in particular, the level of v_m--are derived for differing levels of the strength variable r' (allowing alternative values for r'_o). We seek to derive a new function, which we label a "switch function" for P, for which each point is a switch point, as was (r'_o, v_m). Although we could define the switch function as a function of any subset of $\{r', \ldots, x_m\}$,

we again choose to focus on the relationship between the strength of the case (at trial, r_o') and the minimally acceptable plea necessary to induce P to switch from obeying the trial constraint to obeying the bargaining one. The resulting function can be interpreted to be a demand function for one good--a plea of guilty at level v_m--as a function of the "price" of a substitute good, outcome obtained via trial.[35]

Deriving the switch function point-by-point, we would reason in the following manner: a lower value of r_o' would imply a higher level of U_o; that in turn would require a lower value of r_m' and result in a higher value of v_m. We would thus expect that as trial strength of the prosecutor vis-a-vis the defense falls, the outcome in terms of v would increase also. Then to the new function $v_m = S_p(r')$ we attribute a negative first derivative, but cannot determine the sign of the second derivative without knowledge of the forms of the utility function and trial-and-plea opportunity sets.[36]

Following is an example in which a prosecutor's switch function is derived from a Cobb-Douglas utility function, and simple production function for evidence and resource constraint. All were chosen for use here because they possess the properties required by the theory, yet are simple enough to survive the necessary mathematical manipulations.

Suppose a prosecutor has, or pledges himself to

(i) a utility function $U(v,x) = v^{\alpha}x^{\beta}$, where $0 \leq \alpha, \beta \leq 1$;

(ii) production functions for v and x:

$$v = \frac{1}{r'} \cdot M_v \qquad \text{and} \qquad x = \frac{1}{k} \cdot M_x$$

where M_v and M_x denote the amounts of variable input M which is a mush of all variable inputs to the production by P of a strong case;

(iii) a total disposable resource limitation of R if V is resolved via trial, or R' if by plea of guilty. At a price of w per unit of M, this allows the total variable input to x and v to be $\underline{M} = R/w$ for trial, or $\underline{M}' = R'/w$ for pre-trial settlement.

First it is necessary to derive the production possibility frontiers which constitute the trial and plea opportunity sets. From (iii) we have that $M_v + M_x = \underline{M}$ for trial and $M_v + M_x = \underline{M}'$ for a plea. Since derivation of G is precisely parallel to the derivation of F, only the latter will be performed fully here. Rewriting the trial resource limitation as $M_v = \underline{M} - M_x$, and substituting into the production function, we obtain $M_x = \underline{M} - r'v$. Substituting that expression into the production function for x yields

$$r'v + kx = \underline{M}$$

for the function F. Similarly, G can be written

$$r_p'.v + kx = \underline{M}'.$$

Next, solving the utility maximization problem for U(v,x) subject to F yields

$$v = \frac{\alpha \underline{M}}{(\alpha+\beta)\, r'} \quad , \quad x = \frac{\beta \underline{M}}{(\alpha+\beta)\, k} \quad ,$$

and consequent utility level

$$U_o = \left(\frac{M}{\alpha+\beta}\right)^{\alpha+\beta} \cdot \left(\frac{\alpha}{r'_o}\right)^{\alpha} \cdot \left(\frac{\beta}{k}\right)^{\beta}.$$

The problem then reduces to one of solving the following three equations:

(a) $r'_m v + kx = M'$

(b) $x = U_o^{1/\beta} v^{-\alpha/\beta}$ (which is another way of writing U(v,x))

and (c) $\frac{r'_m}{k} = -\frac{dx}{dv}$, where x(v) is the indifference curve of equation (b). This last equation may be rewritten

$$r'_m = k \cdot \frac{\alpha}{\beta} \cdot U_o^{1/\beta} v^{-\alpha/\beta - 1} .$$

Substituting from (b) and (c) into (a), we obtain

$$v = \frac{M'}{k\left(\frac{\alpha}{\beta} + 1\right)}^{-\beta/\alpha} \cdot U_o^{1/\alpha}$$

and further substituting for U_o the value obtained earlier, as well as dollar resources for units of M,

$$v_m = \frac{(R/w)^{(\alpha+\beta)/\alpha}}{(R'/w)^{\beta/\alpha}} \cdot \frac{\alpha}{\alpha+\beta} \cdot (r')^{-1} .$$

As the reader can verify, the utility function meets the conditions set forth in the theory, i.e.

$$\partial U/\partial v > 0 \quad , \qquad \partial^2 U/\partial v^2 < 0$$

$$\partial U/\partial x > 0 \quad , \text{ and } \quad \partial^2 U/\partial x^2 < 0.$$

Further, the linear opportunity sets are nonconvex with respect to the origin, and, since $\underline{M}' > \underline{M}$, the plea constraint dominates the trial constraint for small values of v.

The switch function which results is also as suggested by the theory: the level of the guilty plea is inversely related to the defendant's strength ($\partial v_m/\partial r' < 0$); for R' fixed, v_m increases with R. The latter reflects not only the general desirability of having disposable resources, but also the effect of pre-trial delay, which serves to reduce R. Finally, if the trial date (and hence R) is held constant, and the bargaining, or plea date is allowed to vary, large values of R' may be undesirable to P; large values of R' in relation to R reflect the need for reallocation of time between the queues so that bargaining occurs closer to the date of the potential conflict. This last statement is consistent with the catalytic effect of impending trial deadlines.

NOTES TO 2.2

[1] Here, the prosecution is assumed to be one unit and is referred to as if it were one person in spite of the fact that many people are involved in the processing of the case and the various decisions made along the way. As with industrial relations and bargaining, a natural question is whether the various members of each unit have coincident or divergent vested interests, and the answer is probably that there exists some conflict. In particular, the interest of the prosecutor in obtaining convictions--his rate on those will become a campaign issue--may conflict with other public goals, such as minimizing personal injury incidence in a certain part of the city. The various potential conflicts are outlined in Lester Thurow's discussion of "Equity Versus Efficiency in Law Enforcement" (1970).

Nevertheless, at least within the prosecutor's office, it is reasonable to conceive of the prosecutor as a single entity, for the court deals with it as such. First, Justice Burger for the Court, then Justice Douglas concurring:

> "...[A]t this stage the prosecution is not in a good position to argue that its inadvertant breach of agreement is immaterial. The staff lawyers in a prosecutor's office have the burden of 'letting the left hand know what the right hand is doing' or has done. That the breach of agreement was inadvertant does not lessen its impact."
> "The staff of the prosecution is a unit and each member must be presumed to know the commitments made by any other member." Santobello v. New York 404 U.S. 257 (1971).

[2] An example of the decision made by default is provided by the execution of an ill-fated "sixty-day law" in Florida. Passed several years ago in the wake of the death of a juvenile incarcerated while awaiting trial, the new law required that defendants not tried within sixty days of their arrest be released. After enaction of the law, the courts apparently continued along on a first-come-first-served basis. When the first group of defendants was released under the law, the public outcry was great, as many accused felons prepared to join them on the streets; the law was quickly made "inoperative."

[3] Theoretically, there is also a problem of time-sequencing in the disposition of the cases relative to one another: while one case is being resolved under the prosecution's attention, others deteriorate as evidence becomes harder to preserve and to find as time passes. Because the deterioration may be uneven from one case to another, each ordering of the cases in time may constitute a choice among goods which are in fact different goods. It has been suggested to me that this is

analogous to the problem of the physicist who finds that the substance with which he is dealing (particularly in nuclear physics) is itself changing through time. However, the time-sequencing problem is ignored here for the present time.

[4] The units of measure for case outcome may be given several representations; for example, the number of years of prison sentence, or a general index of loss of prestige, fines, probation conditions and prison sentence.

[5] In textbooks one is able to find definitions of fixed costs which could yield contradictory classifications of certain kinds of costs. Here, if we interpret output to mean efforts at improving case outcome, some of the costs we have labeled as semi-fixed ones are those which do not vary with output, but do vary with time.

[6] In U.S.v. Wiley (1960), Chief Judge Campbell outlined some of the costs of a case up until the time of sentencing:

> For example, in the case before me, the official total cost to the government...was $11,250. The Assistant United States Attorney, upon representation spent three days in preparation for Wiley's trial and one day on trial. Several investigators likewise spent three days in preparation for trial and one day at the trial. I myself spent one half day in preparation for trial and then one day on trial and accordingly the Court was totally occupied for a day and a half. Though the expense and time that would have been saved the government if Wiley had pled guilty is fairly obvious, I suggest that if this had been a criminal trial lasting weeks or months as many of our trials actually do, then the saving of time and expense to the government would have been far more striking.

[7] For the moment it is assumed that all cases do come to trial. This assumption is discussed further below.

[8] This assumption is made here in order to simplify the analysis, and is relaxed later in the section, when the alternatives of dropping cases (or recommending dismissal) and pleas of guilty are introduced as possibilities.

[9] Each of these possibilities has its counterpart in familiar discussions of the "crisis in the courts" as some like to call it. Politicians and others have called for increased court funding, cutting back the "fat" in the bureaucracy, reducing delay ("Justice delayed is justice denied," as the adage goes.) The last two cost-cutters represent areas of reform which are "hot" and have been exposed to recent experimentation. Although the cost of pretrial detention is frequently

part of the budget of an institution other than the prosecution, the resources thus spent represent a net public expenditure, and subject to accounting limitations, resources saved by reduction of pretrial detention (either a reduction in the number of defendants in jail, or in the period of incarceration) should make possible an increase in the disposable resources of the prosecutor. Some cities and states have changed their record-keeping to systems which not only allow lower maintenance costs, but also permit greater coordination among the several law enforcement agencies. Finally, many governmental units are making proposals for change which would reduce the number of "false alarm" trials, at which witnesses, attorneys, juries, and judges must appear and wait (and be paid) only to find that trial is postponed for one reason or another; the effect of such proposals is to reduce T_v and T_x.

[10] In general, when P faces what he considers to be a tough case, he would expect a worse outcome, ceteris paribus, than he would anticipate from an easy one, i.e. $\partial v/\partial r' < 0$. However, it should be noted that although incantations of ceteris paribus evoke the above result on the opportunity set F, they will not necessarily yield that same result when applied to the solution of the utility-maximization problem, as should become clearer once P's choice mechanism is presented below.

[11] Note that any revisions by P of this estimate will change the benchmark by which he judges the strength of his opponent.

[12] See McIntyre (1965) and Newman (1966) for discussion. The prosecutor in Detroit confirmed that they do this; in fact, a blank labeled "Probability of Winning" appears on the case file forms.

[13] Certain pieces of information--the race, approximate age, sex, custody, social status--would be known by a jury in any event if D went to trial. Others--previous arrests and employment status, for example--could only be brought out by questioning the defendant himself. It is often helpful to D for him to take the stand in his own behalf, but in doing so he raises the possibility of other damaging facts becoming known.

[14] According to Phillip Schaenman of the Urban Institute, a study being conducted there indicated that little sleuthing by police (for themselves or for the prosecution) took place after arrest, although occasionally there might be some additional work done on the eve of trial. For the most part, the evidence gathered around the time of arrest was the total evidence available to P for trial.

[15] The theory of plea bargaining presented here has little to say about the form of this function, and thus can be no more specific than to include a description of the directional effects of each of the contributing variables. For example, we could require the d_i to be stated

(i.e. coded) so that $\partial r/\partial d_i < 0$ for all i.

[16] For example, if F is the function, $v^2 + (x-r')^2 = (x_o-r')^2$, $x, v, r' \geq 0$ and r' restricted to be less than x_0, the opportunity set is an arc of a circle with radius (x_o-r') and center at (0,r'). As r' increases, F will rotate clockwise to F', as illustrated in the text.

[17] This characterization of the opportunity set as the result of a production process is not new in its application to micro-analysis. Gary Becker (1965) developed such an opportunity set to describe the set of choices of a consumer in allocating his resources between market-intensive and (consumer-) time-intensive goods. It is also curious but disappointing to note that although Harsanyi conceived of a production relation underlying a bargaining relationship, he failed to carry the notion any further in his theory:

> Power in a schedule sense can be regarded as a 'production function' describing how a given individual can 'transform' different amounts of his resources (of his working time, his money, his political rights, his popularity, etc.) into social power of various dimensions (of various strengths, scopes, amounts and extensions). (1962)

[18] Unless v or x is an inferior good, the effects of a decrease in r or increase in R are as noted in the text. It is highly unlikely that x (the rest of the caseload, taken as a whole) should be regarded as an inferior good. It might, however, be true for individual cases, such as v. If, say, v is deemed not very serious (e.g. a littering offense) and pursuit of more serious cases is out of financial reach perhaps because of lack of access to lab facilities, etc., the effect of an increase in R could be to shift prosecution efforts from the "small stuff" to more serious offenses. As a non-court example of this, the Boston police claim that with only $80 per narcotics officer per month for purchases, it is impossible for them to do a thorough job in policing narcotics traffic (a minimal purchase would be $65) and therefore they concentrate their efforts on other types of offenses. This is in contrast to the situation in New York, where the figure is several thousand dollars per agent per month. ABC television documentary on the Boston drug trade, "The September Factor," April 1974.

[19] Verification of the effects of these factors exists among the literature of sociology and criminology, and is borne out by the data used from Detroit. See section 3.2 and the data appendix.

[20] This effect of the endowment variable highlights the fact that the ability to attain certain case outcomes at trial depends intimately upon the earlier work of the police.

[21] Under certain circumstances the changes in r' and/or R could result in a major shift of resources from x to v (or from v), although such a shift would require special conditions on the utility function. This might happen in a highly publicized case, as the details became known and pressure was exerted on the prosecution to raise the priority of the case.

[22] A hypothetical situation would be one in which two witnesses to a crime later meet, piece together their partial understandings of the incident, and appear on the prosecutor's doorstep.

[23] In the strategy of the defense in nearly every criminal case delay is given a prominent place. It often means absence of witnesses, and in the case of repeated continuances, disgusted and angry witnesses for the state, the memory of the witnesses becomes dimmed, the prosecutor may be changed, he may lose interest as the case grows stale, and in many ways a decided advantage for the defendant in the average case may be had by the simple expedient of putting off the trial.

Warren and Beatty (1932) referring to the Missouri Crime Survey.

[24] Witnesses die, forget, or become unavailable; the heroin in the French connection case being stolen are examples of this.

[25] The costs of trial relative to those of a plea of guilty form a ratio of eight or ten to one. One result of the high proportion of guilty pleas--and of institutions which have adapted themselves to the use of that resource-saving method of disposition-- is a sense of desperation at the suggestion that all cases be given full trials. In a magazine interview (Mills, 1971) New York defense attorney Martin Erdmann described a time when he felt a client was being treated unfairly and threatened to insist on trial for all the clients of the Legal Aid Society where he worked, the latter accounting for the defense of 60 percent of the city's criminal caseload. The prosecution compromised.

[26] The discretionary role derives from the separation of powers doctrine in the U.S. Constitution. See Remington casebook (1969) and supplement (1972) for sequence of cases affirming and explicating prosecutional discretion.

An example of the usefulness of such discretion is given by Katzman (19 : 436) for the police:

> The existence of many people vulnerable to arrest in a neighborhood can be a valuable resource to the policeman who can stock up on 'potential arrest.'...A patrolman may deliberately not detect a petty criminal...who in exchange might provide information about more serious crimes in the future.

[27] The use of negotiated pleas of guilty is a controversial issue. The ABA has taken a position favoring such pleas so long as they are obtained through certain procedures which the Association believes safeguard the rights of the defendant and the interests of the community. (See Standards for Pleas of Guilty, 1967).

Arguments in favor of guilty pleas are the efficiency of the system (assuming the defendant is actually guilty), and the belief that admission of the crime by a guilty defendant indicates a willingness to cooperate and an interest in rehabilitating himself. Opponents argue that the purpose of the trial process is not efficiency, but justice, and encouraging defendants to waive the right to trial does less than insure the latter. They further argue that the agreement to relinquish the right to trial often has the effect of losing other rights as well, such as the Fifth Amendment right not to incriminate oneself. A theory which considered the output of a court to be not only dispositions of cases fed into it, but also the explication and preservation of individual rights, including a right to safety, might be able to accommodate arguments from both sides of this issue.

[28] Alternately, if zero disposable resources were the way of life for prosecutors, the role of and processes employed by prosecutors would in effect determine a different institution from what currently exists in most cities.

[29] Remington and others point out (1969: 603) that

> There are several motivations on the part of the prosecutor and the courts in compromising with the defendant. A major one...is avoiding the cost in time and effort of trying cases unnecessarily. Another reason is the uncertainty of conviction.

[30] There is an unlikely but possible problem here: if the second derivative of G were sufficiently large (sufficiently greater than that of F) and if we were to allow the v-intercept of G to be greater than or equal to that of F, we could encounter two intersections of the opportunity sets, with the result that plea bargains would yield better outcomes for P at the extremes, and the mediocre or medium-low-penalty cases would be better taken to trial. Given P's assumed risk aversion, this would be a surprising conclusion.

[31] Since P cannot regard all of his cases to be inferior goods, the presence of some such cases merely strengthens the above argument as regards the others.

[32] One interesting example is of a plea bargain for reduction of charge from breaking and entering at night to breaking and entering during the day--in which case the time of the offense would not be stressed!

[33] This is consonant with Harsanyi's definition of power in bargaining: "The common-sense notion of social power makes it the ability to achieve certain things--an ability that the person concerned is free to use or to leave unused." (underscoring his) (1962: 87)

[34] As one prosecutor phrased it, this compromise is necessary to indicate that "we're really talking about a plea."

[35] In the sense that either a plea of guilty or a trial satisfies the legal conditions for disposition of a case, and in the sense that neither P nor D desires more than one of these to occur for a given offense, these "goods" are perfect substitutes for one another. However, in terms of characteristics and some of the consequences of pleas of guilty, such as limitations on the ability to appeal, the relationship is not so simple.

[36] From the vantage point of this stage in the development of the thesis, one can see some of the difficulties encountered in attempting to apply theories involving the use of the incremental utility functions. Since the determination of the Pareto optimal interval--so far, we have determined the endpoint contributed by the prosecutor--is dependent upon r', the incremental utility function is in fact a function of two variables, $(\bar{v}-v_m)$ being the obvious one addressed by the theory, and r' being the one "hidden" in v_m.

2.3 The Defense

Unlike the prosecutor, the defendant cannot choose to drop the particular case in which he is involved, but does have a choice not only in whether he pleads guilty or not,[1] but also--whichever way he pleads to the charges against him--to decide how vigorously he will fight the case. He makes this latter decision (which in turn affects the former one, also) based upon his attitudes toward different trial outcomes and the opportunity set which he confronts for trial.

The defendant's utility is a function of the outcome of the case in question and of the amounts of other goods (and services, and savings, etc.) We represent the former again by v, and the latter by c. To this utility function u(v,c) we ascribe the usual properties concerning changes in u as a function of changes in c, i.e.

$$\frac{\partial u}{\partial c} > 0 \qquad \text{and} \qquad \frac{\partial^2 u}{\partial c^2} < 0$$

Clearly, v is undesirable to the defendant, and so we expect $\partial u/\partial v < 0$; the sign of the second derivative is somewhat less obvious, however.

For desirable goods, such as c, our assumption of diminishing marginal utility is equivalent to an assumption of increasing immunity or indifference toward incremental quantities of the good as a function of the amount one already has: an additional unit of c means less to a person who already has a lot than it would were he to have little to begin with. The counterpart assumption for an undesirable good such as v is that the more of v a defendant already has, the less difference an additional unit makes in his utility level. Thus for a person confronted with a one-year sentence, adding another year

would cause a greater utility decline than would an added year if his anticipated sentence were already twenty years. If we accept this premise, we have effectively assumed that $\partial^2 u/\partial v^2 \gtrless 0$--an assumption which, although appropriate in this context, is unusual within traditional economic theory,[2] and holds some rather different implications for outcome than would its opposite.[3] In particular, it is likely that when he subjects this utility function to some opportunity sets/constraints, P will find that in order to maximize his utility, he must choose between two corner solutions; this possibility is only mentioned here to foreshadow results to come much later in this section.

Since it is infrequent that one discusses utility in terms of both desirable goods and undesirable ones[4] a substitution for v is made by defining a new good V^* whose units, v^*, measure improvement in v relative to some benchmark, such as the maximum possible level of v: $v^* = \bar{v} - v$, where $\bar{v}$ is the benchmark amount of V. Thus defined, V^* is a desirable good, that is, $\partial u/\partial v^* > 0$, but retains the positive second derivative from before.[5]

We also require that the indifference curves in v^*-c space have negative first derivatives and have second derivatives which are exclusively negative or positive throughout the domain of v and c.[6] Further, since the case in which the second derivative is positive is precisely the analog of the situation encountered by the prosecutor, the discussion here concerns itself with the case in which the second derivative is negative.[7]

The trial opportunity set confronted by the defendant is similar in its development to its counterpart for the prosecutor, with the

major exception that the level of resources available for allocation between c and v^* is no longer a predetermined amount received from an outside source such as a city council or state government, but rather is dependent upon a number of factors relating to the defendant and to his case. Paramount among these factors are D's source and amount of income, pre-trial jail status, and expected times for trial and for pre-trial bargaining (or other method of handling a plea of guilty.)

For the period up until trial, D's total resources are:

$$R_T = W_o + W_1 t_v + (1-J) \cdot Y \cdot t_v \cdot$$

where R_T represents total resources; W_o his wealth or initial level of resources; W_1 is income that is not dependent upon jail status, such as income from certain transfer programs or from forms of non-human investment per unit time ; Y is D's income which is dependent upon his jail status such as wages, unemployment compensation, or stipend from job training programs per unit time; J is his jail status, with J = 0 if D is out of jail pending trial, and J = 1 if he is in jail; and t_v, as in the preceding section, represents the expected time until trial, where the time of arrest is considered to be zero.

As did the prosecutor, the defendant experiences both costs which are fixed (assuming he goes to trial), which are labeled B_o; and ones which are time-dependent, associated with maintenance of his case, such as attorney fees above the retainer , labeled B_1.

Taking the defendant's disposable resources, $\underline{R}$ to be the difference between his total resources R_T and the costs associated with pursuing his case,

$$\underline{R} = W_o + W_1 t_v + (1-J)Yt_v - B_o - B_1 t_v$$

which can be restated as

$$\underline{R} = W_o - B_o + (W_1 - B_1)t_v + (1-J) \cdot Y \cdot t_v$$

where $\underline{R}$ represents the defendant's disposable resources; B_o and B_1 are the fixed and time-dependent costs of his defense; W_o is his initial resource level; W_1 and Y are his time-dependent sources of income, which are respectively independent of and dependent upon jail status; J is unity if D is in jail and zero otherwise; and t_v is the time from arrest until trial. Although it is possible for some of these values to be zero or less (for example, Y equal to zero in the case of a person living entirely from wealth, or $\underline{R}$ negative for a person in debt) we redefine $\underline{R}$ to be zero in the event that a negative $\underline{R}$ would otherwise result from the above computations in the beliefs that (1) the borrowing capacity of persons who are defendants in criminal trials is usually quite limited; and (2) due to judicial decisions of recent years, indigent defendants have been given the right to assigned counsel, in which case B_o and B_1 should be close to zero if $\underline{R}$ would otherwise be less than or equal to zero.[8]

This representation of the level of D's disposable resources permits easy analysis of the effects of alternative values of each of the variables which comprise $\underline{R}$. For the moment ignoring the evidence effects of delay, clearly a defendant who is employed and out of jail pending trial will have a very different attitude toward delay than would that same defendant if employed and in jail or if receiving no jail status-dependent income (and in jail.) Thus, it is impossible

to state the resource effects of delay upon D without additional knowledge about the origin of his income and about his jail status.

The effects of alternative values for the other contributors to $\underline{R}$ are obvious: Other things being equal, a defendant has greater resources if he has a higher income (of whatever origin, whether W_1 or Y), greater wealth, lower case maintenance costs, and lower fixed trial expenses.[9] One can also see the effects of various subsidies upon the level of D's disposable resources. For example, if the defendant's family is on welfare--although the amount of the payment might decline somewhat if D were in jail and his allotment subtracted from the total--his level of disposable resources would remain largely unaffected by his jail status, assuming he has a negligible amount of other jail status-dependent income.[10] As another example, subsidies to defendants in the form of legal services would cause $\underline{R}$ to increase, as noted above; however, such programs may limit the ability of a defendant, who to qualify had to claim indigence, from contributing any of his resources to the cause of bettering his trial outcome, i.e. he might be forced by institutional rules to allocate all of his resources to c.

Having disposable resources $\underline{R}$, as determined above, the defendant chooses to allocate $\underline{R}$ between c and v^*, and does as subject to his trial opportunity set f (again, this function need not be linear, but is assumed to be nonstrictly concave with respect to the origin) which, as with F, is derived as a production possibility frontier. Underlying f are production functions for v^* and c, but the production relation for c is merely an identity function. Stated in implicit form, the trial opportunity set is $f(v^*,c) = 0$, with

$$-\frac{\partial f/\partial v^*}{\partial f/\partial c} < 0 .$$

The defendant's trial opportunity set depends, in addition to v*, c, and R, upon the relative strength of the parties, this time as viewed by the defendant, which is here labeled r". As was done earlier with the prosecutor, r" is defined to be large when the opposition is stronger than for the lower values of r". Note, however, that a stronger case for D is associated now with low values of r" and high values of r'. Like the prosecutor, the defendant makes an assessment of his own initial strength and his enthusiasm for exploiting its possibilities; consequently, alternative values of r" are reflective of alternative estimates of P's strength by D.

The location of f changes in response to differing possible values of r" in the same manner as did F due to alternative values for r' in the earlier analysis of P's opportunity sets. These changes in r" are precipitated by differing states of the same factors which affected r', namely the initial state of the evidence, and characteristics of the defendant and of the offense (although the resulting change in r" is in the opposite direction from its corresponding effect on r'.) The weaker is P's perceived strength, the lower is r"; if we compare f at one value of r" with the same function f' based upon a lower value, the contrast would be as sketched below:

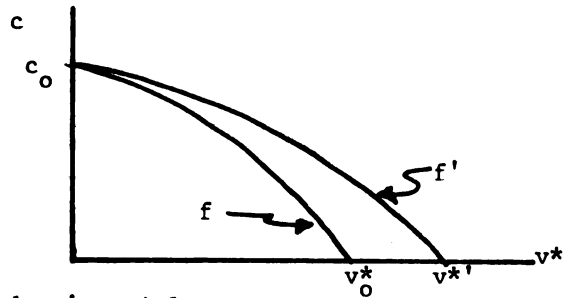


Figure 11. Defendant's trial constraint under alternative values of r".

The defendant seeks to maximize his utility subject to the opportunity set f; however, depending upon the specific trial constraint and utility function, D may be able to do this only by specializing in one good or the other:

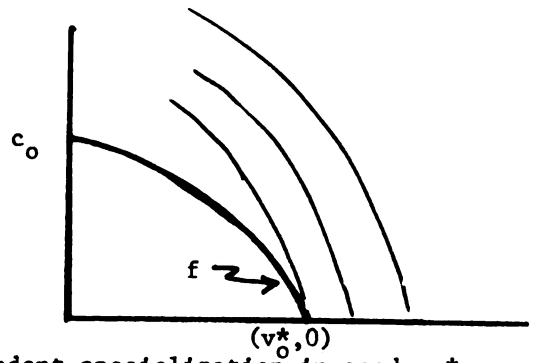


Figure 12. Defendant specialization in goods v* or c.

Were the second derivative of the opportunity set $(d^2c/dv^2)_f$ to be greater than that of the indifference curves--graphically, the constraint would have a "tighter" curvature than would the indifference curves--such specialization would not be necessary, and the maximization problem would revert to one sketched below, which is conceptually the same as the situation described in the preceding section for the prosecutor

(and hence that case is not pursued here.)

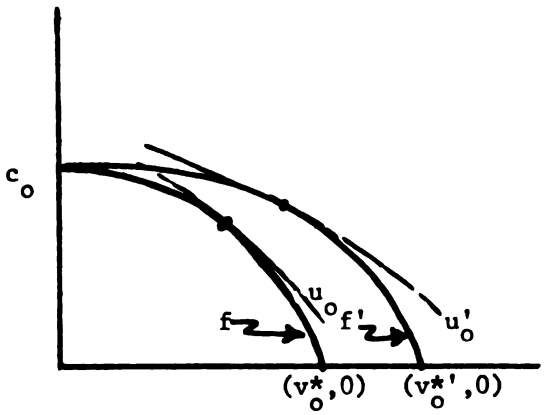


Figure 13. Defendant non-specialization.

However, if the second derivative of f is less than that of the indifference curves, a search for extrema other than at the boundaries would yield a point of minimum, rather than maximum, utility for D. If this latter situation obtains--for example, in the event that f is linear--D will choose the specialization that yields the higher level of utility. Let $u_o^1 = u(0,c_o)$ and $u_o^2 = u(v_o^*,0)$. Then $u_o = \max\left\{u_o^1, u_o^2\right\}$, and u_o is the utility D associates with trial. Examples of situations which would give rise to specialization in v* or c, respectively, are drawn below.

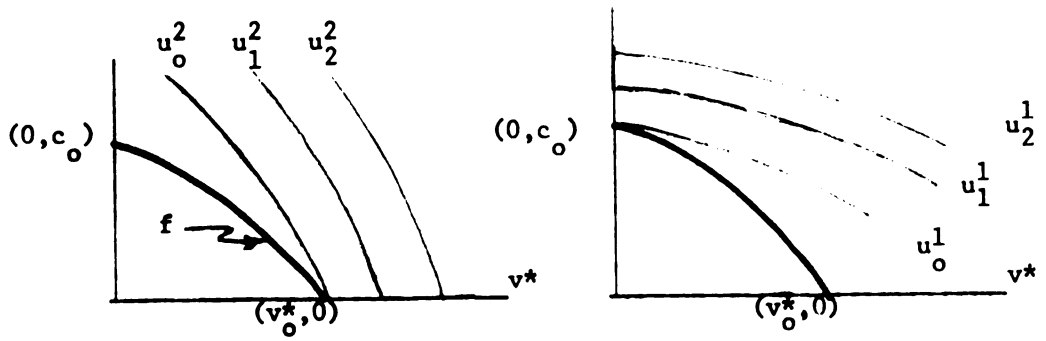


Figure 14. Defendant specialization in case improvement; in consumption.

Clearly, the value of r" may be critical in determining the defendant's approach to dealing with his case; for the first example above, by choosing successively larger values of r", could be transformed into the situation depicted by the second example, thus causing D to reverse his specialization, even though all other factors have remained unchanged. Thus, differing values of components of r", in particular values of e, the stock of evidence, could result in vastly differing outcomes v^*, even though the difference in the r"-values might be very small. [11]

Allowing r" to vary between values such as it took on in the examples above, we can determine the special values r_s'' of r" such that

$u_o = u_o^2$ for all $r'' < r_s''$ (specialization in v^*)

and $u_o = u_o^1$ for all $r'' > r_s''$ (specialization in c.)

Then for $r'' = r_s''$, $u_o = u_o^1 = u_o^2$, and either mode of specialization will yield the utility-maximizing result. [12]

Having considered his various outcomes to be obtained under the assumption that he observes the trial opportunity set f, the defendant also makes similar judgements concerning an alternative opportunity set, which represents the set of possible outcomes if he should plead guilty. [13] Pleading guilty offers a number of possible advantages to D: in addition to reducing the maximum sentence he could receive, and to avoiding some of the undesirable publicity and psychic costs of trial, the defendant is able to conserve some of the resources which would otherwise have been spent on the costs of case maintenance (for the post-plea bargaining but pre-trial period) as well as the costs

associated with the trial itself. Thus the level of disposable resources on the guilty plea opportunity set g is

$$\underline{R}' = W_o + (W_1 - B_1)\, t_p + (1-J)\, Y\, t_p \ ,$$

where t_p is the time from arrest to the entering of a plea of guilty. The difference between $\underline{R}$ and $\underline{R}'$ is then

$$\underline{R} - \underline{R}' = B_o + (W_1 - B_1)\,(t_v - t_p) + (1 - J)\, Y\, (t_v - t_p) \ ,$$

which we expect will be greater than or equal to zero, implying that g will have a larger c-intercept than does f.

The defendant performs the same exercise of utility maximization subject to g that he executed when f was being considered as the constraint. Since, (for the same reasons we included the restriction on the prosecutor) we expect the v*-intercept of g to be less than or equal to the intercept for f, we concern ourselves here only with the utility value associated with $(0,c_o')$, in the expectation that $u_o' = \max\{u_o'^1, u_o'^2\} = u_o'^1$, [14] where $u_o'^1 = u(0,c_o')$ and $u_o'^2 = u(v_o^{*\prime},0)$. The graph of the two opportunity sets which D confronts is the same picture as that of P's opportunity sets, save for the labels. Below is the graph for D, with the utility levels indicated for each intercept.

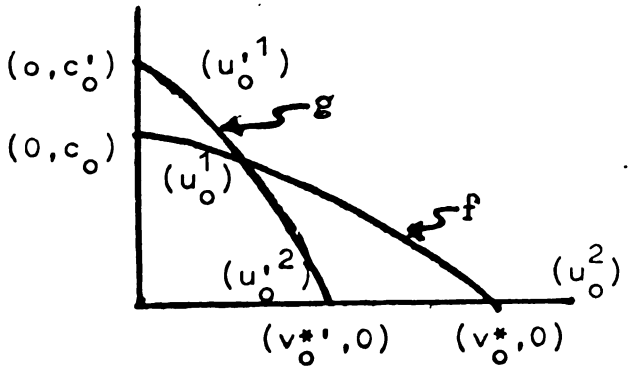


Figure 15. Defendant's trial and plea opportunity sets.

After determining the utility levels obtainable at trial and by plea of guilty, u_o and u_o' respectively, the defendant chooses the option which affords him the greater utility level. However, as before, if $u_o' < u_o$, appropriate initial concessions by P again effected by failure to exercise the full strength of his case, could make a plea of guilty sufficiently attractive to induce the defendant to so plead. In taking a closer look at D's comparison of trial outcome with that of a plea, and at the determination of the concession necessary should the latter opportunity set not be an instant success, three cases are considered: (1) $r_o'' > r_s''$, the special level of r" on f at which both trial and consumption-specialization would result in the same utility to D; (2) $r_o'' < r_s''$ and $u_o'^1 > u_o^2$; and (3) $r_o'' < r_s''$ and $u_o'^1 < u_o^2$. These are shown together in the figure below; then each is analyzed in the pages which follow. As can be discerned from the sketches and from the discussion, varying the values of r" and c_o' relative to c_o is sufficient to transform one case into another.

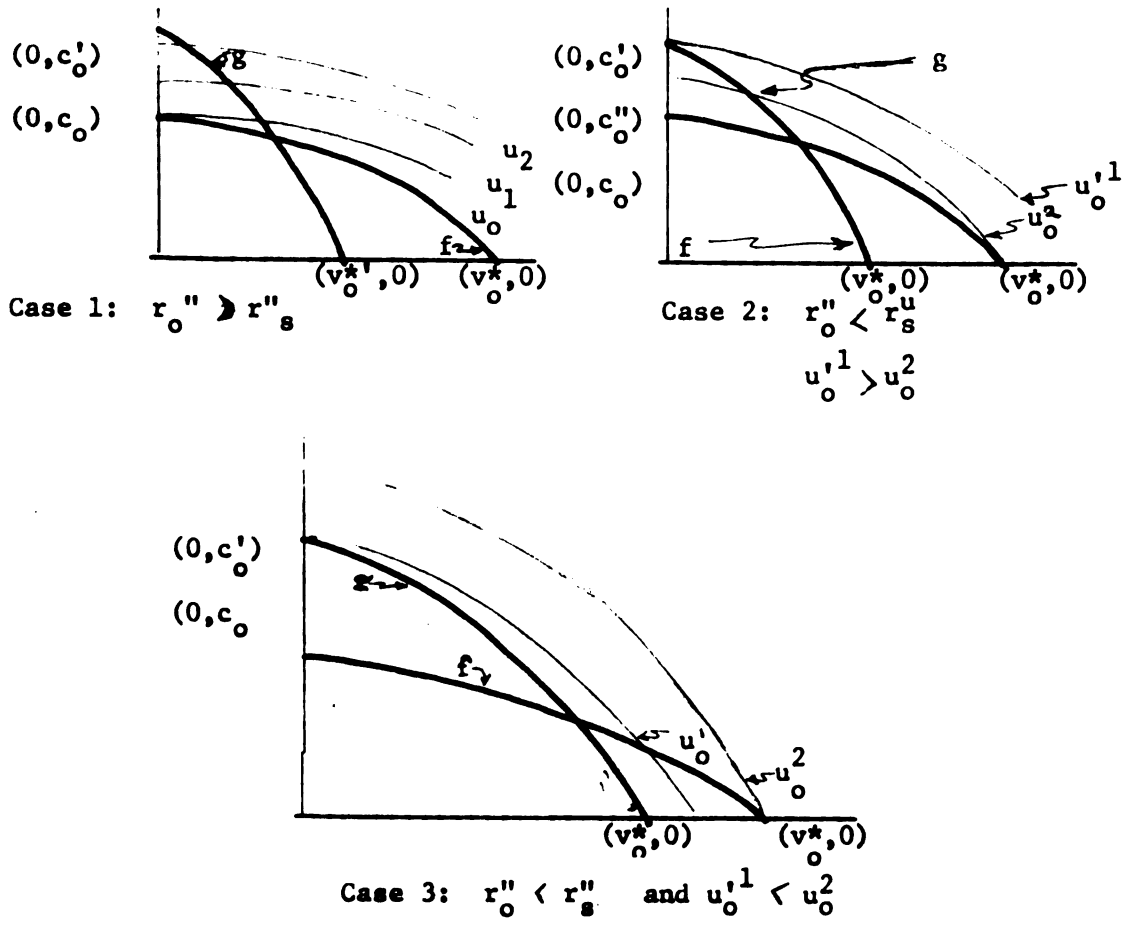


Figure 16. Possible relationships of trial and guilty plea opportunity sets.

In the first of these situations, the defendant chooses to specialize in c if obeying the trial constraint (since $r''_o \rangle r''_s$) and so the availability of the guilty plea option merely offers him the opportunity of being able to obtain more c--with the same v-outcome--than if he went to trial. [15]

In the second situation also, when $r''_o \langle r''_s$ and when obtaining the utility gain possible from pleading guilty would cause a reversal in

D's specialization from v* to c, the defendant is clearly better off by observing opportunity set g. In both this and the previously described situations, no concession by the prosecutor is necessary in order to make g attractive to the defendant: in the first case, D had not intended to exert much (resource) effort toward improving his case in any event, and so any resource saving made possible by avoiding trial represents pure gain to him; and in the second, the resource saving more than compensates D (in terms of utility units) for the less desirable outcome in terms of v*.

It is in the third situation that a plea of guilty is not immediately preferable to going to trial, and in which some--in fact, perhaps quite a large--concession from P is necessary in order to bring acceptance of a guilty plea into the realm of the possible. Since in this case $r''_o < r''_s$, the defendant is specializing in trial improvement (V*) on f, and expects outcome v^*_o to be the result. Since $u'^1_o < u^2_o$, the mere fact of the resource saving possible on g relative to f is not sufficient to pose g as a viable alternative to the latter, and concession by P would be necessary to change D's current preference for f. Since specialization will take place in either event on both constraints, P must concede to D an outcome set which is at least as good as the outcome D expects to obtain at trial. This compromise is accomplished by P's promise not to exercise his full strength r''_o, but rather to exercise some lesser amount $r''_{promise}$. In effect, the compromise necessary--with the level of r" utilized by P being r''_m--that is the one which causes the v*-intercept of g to be equal to that of f, namely v^*_o. [16] It is important to note that v^*_o does not necessarily denote an outcome of

zero, when stated in terms of v, the undesirable good. As originally defined, v_o^* was simply the best trial outcome that could be attained via f, and hence an offer of outcome v_o^* by P may in fact be an offer which does not let D off "scot free". [17]

Searching for the magic value of r" at which D is indifferent between observing f or observing g leads to r_m'', the value of r" for which g passes through the point $(v_o^*,0)$; that is, the value of r" which results in the same outcome (not just the same utility level) as the trial specialization on f, illustrated below.

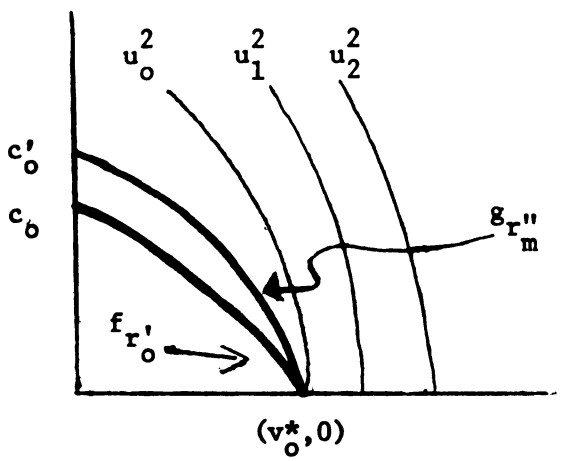


Figure 17. The adjusted guilty plea constraint.

Once that value of r_m'' has been promised, i.e., once $r''_{promise}$ is less than or equal to r_m'', D will be willing to consider pleading guilty at $v_m^* = v_o^*$. Since this was also the outcome expected if D went to trial, we can say that for cases in this third class, the outcome will be v_o^*; for $r''_{promise} \leq r_m''$; this outcome can be attained by a plea of guilty, and otherwise by trial. [18] Thus the switch function for D is identical to the normal demand function for situations in the third class.

It is yet to be shown the manner in which the three cases dissected above descend into one general situation from which two key parameters are permitted to assume alternative values. To do so, we return to the third case and recall that originally, i.e. prior to any concession by P, u_o was greater than u'_o, where both depended upon the original value of r", r''_o. That is, for the given value of r''_o, D preferred to observe the trial constraint and to specialize in trial outcome improvement while doing so. Now let r''_o increase until u_o falls to u'_o, represented graphically by rotating f in a clockwise manner until the v-intercept of f meets the same indifference curve as the c-intercept of g (the latter has remained constant.)

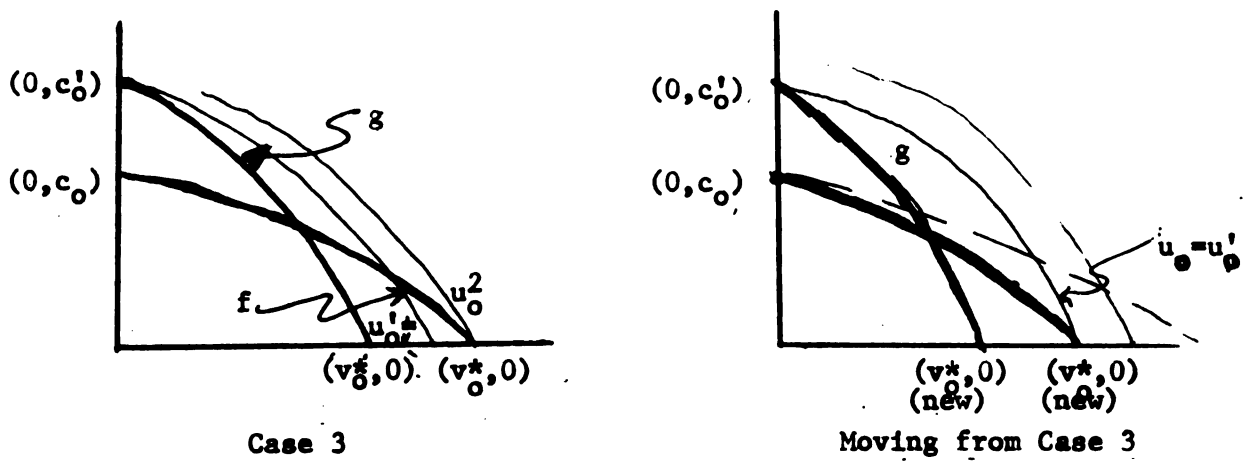


Figure 18. Case 3 and movement from Case 3 by increasing r''_o.

This can be accomplished algebraically by solving the equation of the indifference curve at utility level u'_o for its v*-intercept, then solving the equation of f for the value of r" resulting in that same v*-intercept. Labeling this value of r" by $r''_{disc.}$ we can see that as r" assumes values larger than $r''_{disc.}$, Case 3 no longer obtains, but rather, Case 2.[19]

Allowing r" to take on still higher values, we arrive at Case 1. The sequence of these moves is sketched below.

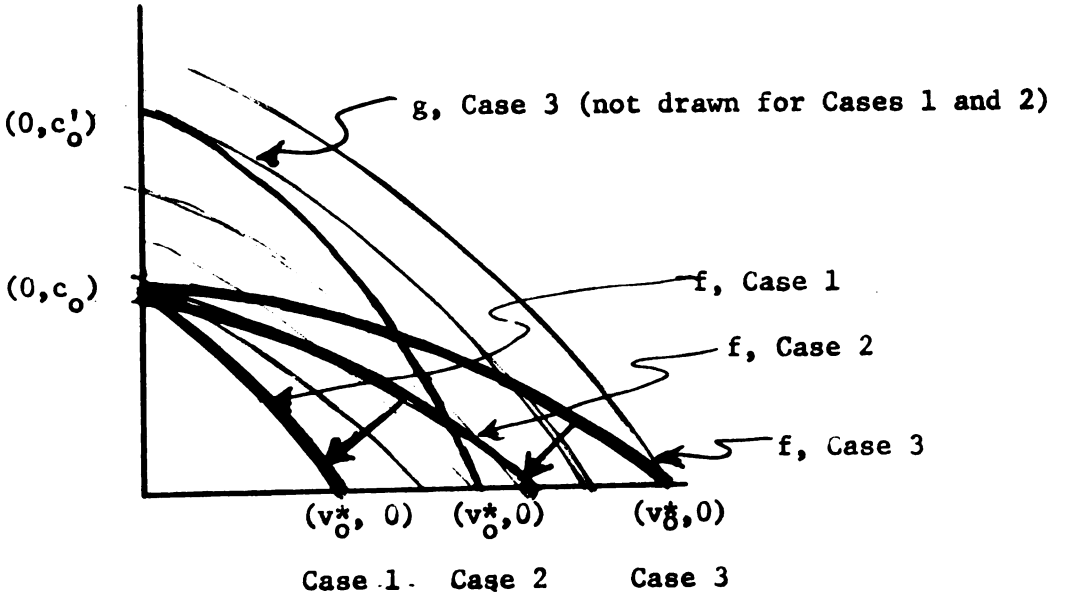


Figure 19. Increases in r''_o lead from Case 3 to Case 2 to Case 1 (Indicated by clockwise-pointing arrows.)

As can be seen from the preceding figure, D would engage in trial improvement specialization in Cases 3 and 2. However, in Case 2, that preference on the f constraint is eclipsed by his switch to opportunity set g, and hence consumption specialization on the latter constraint. As r" is further increased and Case 1 is the relevant situation, the preference now on f for consumption specialization is again overshadowed by the dominance of g: this is the case in which the availability of the guilty plea option does not reverse D's specialization, but merely increases his disposable resources.

To summarize the conclusions from the case-by-case analysis: In Case 3, D is initially specializing in trial outcome improvement on f and plans to have the charges against him disposed of at trial. If the

prosecutor is to interest D in pleading guilty, P will not only have to make some concession, but will have to offer D a charge at least as low as what D otherwise expects to obtain from trial. Then for each level of r''_o, $v^*_m = v^*_o$, and the switch function (in Case 1) traces out the ordinary "demand" function for v^*_o in terms of r''_o, i.e.

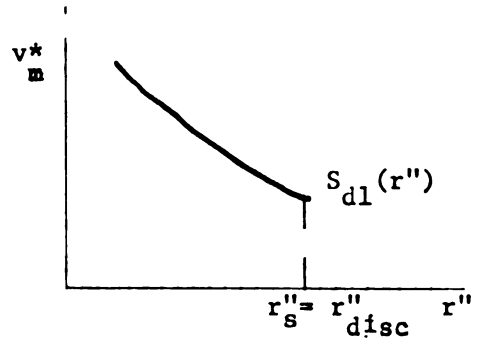


Figure 20. The defendant's switch function in terms of v^*_m, for $r'' \leq r''_{disc}$.

where the r''_o and v^*_o for each point of the switch function correspond to the r" parameter and v*-intercept of a trial opportunity set.

Further, since the other two cases resulted in consumption specialization on g, $v^* = 0$ for both. Since Cases 1 and 2 both entailed values of r''_o less than $r''_s = r''_{disc.}$, we have the right-hand part of the defendant's switch function S_{d2} uniformly zero (i.e. $v^* = 0$) for $r'' > r''_{disc}$.

The switch function for the defendant is then $v^*_m = 0$ for $r''_o > \min\{r''_1, r''_s\}$ and $v^*_m = v^*_o$ otherwise, i.e. for $r'' < \min\{r''_1, r''_s\}$. For the lower half-domain, the switch function is identical to the demand for v^*_m as a function of r''_o, and so we know that $dv^*_m/dr'' < 0$;

for the upper half-domain, the switch function is a straight line.

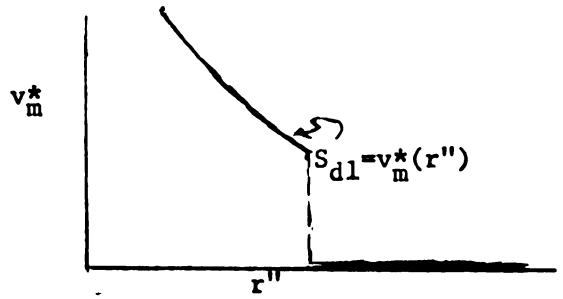


Figure 21. The defendent's switch function, with v_m^* a function of r".

Re-transforming v* back into v (to facilitate combining the theory developed here with that concerning the prosecutor), we write D's switch function as

$v_m = S_d(r'') = v - v_m^*(r'')$, sketched below.

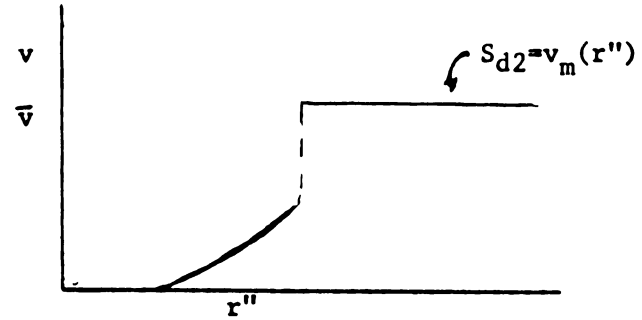


Figure 22. The defendent's switch function in v_m as a function of r".

In the next few pages, an example of the defendent's switch function is derived for a particular utility function and linear constraint. It was not only necessary for the utility function to exhibit the properties required by the theory -- such as appropriate signs of partial derivatives -- but also that it be manipulable throughout the steps required for the derivation, and that the final function attained not require statement in implicit form. The result

of considerable searching, the following function was the simplest one found which met the above criteria. Let D have the utility function,

$$u(v,c) = (v+1)^{-1} + \ln(c+1)$$

satisfying the conditions that

$$\frac{\partial u}{\partial v} < 0 , \qquad \frac{\partial u}{\partial c} > 0$$

$$\frac{\partial^2 u}{\partial v^2} > 0 , \qquad \frac{\partial^2 u}{\partial c^2} < 0$$

and

$$\frac{dc}{dv} > 0 , \qquad \frac{d^2 c}{dv^2} < 0 .$$

And suppose also that he confronts

$$p_c c + r''(\bar{v} - v) = \underline{R}'$$

and

$$p_c c + r''_{promise}(\bar{v} - v) = \underline{R}'$$

as trial and guilty plea opportunity sets, respectively, where these opportunity sets can be derived from production functions of the form used in the example of section 2.2.

For a first step, it is necessary to state u_o^1, u_o^2, and $u_o'^1$ for the given utility function and constraints. Since the model provides that D always specialize in v-reduction or in C, it is necessary only to solve for the corner solutions and for the point of discontinuity which severs Case 3 from Cases 1 and 2.

Recall that for values of r" falling into Case 3, the switch function was the set of ordered pairs (r''_o, v_o). Substituting from the trial constraint,

$$v_o = \bar{v} - \underline{R}/r''_o \quad \text{(assuming } P_c = 1\text{)}$$

and so, allowing alternative values of v_o and r''_o,

$$S_{d1} = v(r'') = \bar{v} - \underline{R}/r''$$

for $r'' \leq r''_{disc}$.

On the trial constraint, the utility derived from specialization in c is

$$u_o^1 = u(\bar{v}, c_o) = (\bar{v}+1)^{-1} + \ln(c_o+1)$$

and from specialization in v-improvement is

$$u_o^2 = u(v_o, 0) = (v_o+1)^{-1}$$

where v_o is the intercept of f. Substituting from f into u_o^1 and u_o^2 the expressions

$$c_o = \underline{R} \quad \text{and} \quad v_o = \bar{v} - \underline{R}/r''$$

(assuming $p_c = 1$) we have

$$u_o^1 = (\bar{v}+1)^{-1} + \ln(\underline{R}+1)$$

and

$$u_o^2 = (\bar{v} - \underline{R}/r'' + 1)^{-1}$$

Similarly, on the guilty plea constraint, specialization in c yields utility level

$$u_o'^1 = u(\bar{v}, c_o') = (\bar{v}+1)^{-1} + \ln(c_o'+1).$$

Substituting from g into $u_o'^1$,

$$u_o'^1 = (\bar{v}+1)^{-1} + \ln(\underline{R}'+1),$$

to use the notation from the theory.

For $r'' > r''_{disc}$, the defendant is specializing in consumption and $v = \bar{v}$, since trial outcome is surrendered completely. Thus $S_{d2} = v(r'') = \bar{v}$ for $r'' > r''_{disc}$.

Combining these results, the defendants switch function is

$$S_d = \begin{cases} S_{d1} = v(r'') = \bar{v} - \underline{R}/r'' & \text{for } r'' \leq r''_{disc} \\ S_{d2} = v(r'') = \bar{v} & \text{for } r'' > r''_{disc} \end{cases}$$

It remains for us to determine the value of r''_{disc}. As defined

earlier, $r''_{disc} = r''_s$ was the special value of r" at which D was indifferent (on the trial constraint) between specializing in trial outcome improvement or in consumption. To determine r''_{disc} we set u'_o equal to u^2_o and solve for r", which is the value we seek. The requirement that $u'_o = u^2_o$ is equivalently stated

$$\frac{1}{\bar{v}-\underline{R}/r''_{disc}+1} + \ln(c+1) = \frac{1}{\bar{v}+1} + \ln(\underline{R}'+1)$$

where the value of v_o corresponding to r''_s is $(\bar{v} - \underline{R}/r''_s)$ on the g constraint. Combining terms and solving for r''_s,

$$\bar{v} - \frac{\underline{R}}{r''_{disc}} + 1 = \frac{1}{(\bar{v}+1)^{-1} + \ln(\underline{R}'+1)}$$

so

$$\frac{1}{\bar{v}-\underline{R}/r''_{disc}+1} - \frac{1}{\bar{v}+1} = \ln(\underline{R}'+1),$$

and

$$\frac{(\bar{v}+1) - (\bar{v}-\underline{R}/r''_{disc}+1)}{(\bar{v}+1)\ (\bar{v}-\underline{R}/r''_{disc}+1)} = \ln(R'+1) .$$

Combining terms in the numerator of the lefthand side of the equation,

$$\frac{\underline{R}/r''_{disc}}{(\bar{v}+1)\ (\bar{v}+1-\underline{R}/r''_{disc})} = \ln(\underline{R}'+1) .$$

Then $$\frac{\underline{R}/r''_{disc}}{(\bar{v}+1)\ \ln(\underline{R}'+1)} = \bar{v}+1 - \underline{R}/r''_{disc}$$

so that $$\frac{\underline{R}}{(\bar{v}+1)\ \ln(\underline{R}'+1)} = r''_{disc}\ (\bar{v}+1 - \underline{R}/r''_{disc})$$

$$= r''_{disc}\bar{v} + r''_{disc} - \underline{R}$$

and $$\left(\frac{1}{(\bar{v}+1)\ \ln(\underline{R}'+1)} + 1\right)R \quad = \quad r''_{disc}\ (\bar{v}+1)\ .$$

Combining terms on the left and dividing by $(\bar{v}+1)$,

$$r''_{disc} = \frac{1+(v+1)\ \ln(\underline{R}'+1)}{(\bar{v}+1)^2\ \ln(\underline{R}'+1)}\ .\ \underline{R}\ .$$

In terms of the graph notation of the theory

$$r''_{disc} = \frac{1\ +\ (\bar{v}+1)\ \ln(c_o'+1)}{(\bar{v}+1)^2\ \ln(c_o'+1)}\ .\ c_o$$

This gives us the point of discontinuity $r''_s = r''_{disc}$.

The switch function of the defendant has been found to be

$$S_d = \begin{cases} \bar{v} - \underline{R}/r'' & \text{for } r'' \leq r''_{disc} \\ \bar{v} & \text{for } r'' > r''_{disc} \end{cases}$$

where $$r''_{disc} = \frac{1\ +\ (\bar{v}+1)\ \ln(c_o'+1)}{(\bar{v}+1)^2\ \ln(c_o'+1)}\ .\ c_o\ .$$

Again, as with the example given for the prosecutor, the characteristics of the model can be verified here. Clearly, when r" is large, i.e. when the prosecutor's strength is great, the outcome will be large also. When the defendant's resources are great, r''_{disc} is larger, and the non-surrender outcomes comprise more of the set of possible outcomes, thus bettering the plea bargain outcomes for at least some values of the independent variable representing relative strength.

Joining the theory of the defendant with that of the prosecutor, the outcomes of plea bargains can be circumscribed and, subject to imposition of a bargaining theory, predicted, as is done in the section to follow.

NOTES to 2.3

[1]His right to trial is guaranteed by the Sixth Amendment. Although it is most always possible for a defendant to plead guilty, this act must meet the conditions of being intelligent and voluntary, as determined by the judge. And although it is typical that the plea be made upon either the promise or upon the expectation of some sort of leniency in some subsequent stage of the criminal justice process (such as sentencing), such leniency is not guaranteed. See Remington (1969). Finally, the rights of the defendant to have his plea of guilty accepted and honored--given that the plea is intelligent and voluntary--have been affirmed by the courts. McCoy (1966), Santobello (1971).

[2]The case in which $\partial^2 u/\partial v^2 < 0$ is explored in an appendix. The results are analogous to those for the prosecutor, which we derived earlier.

[3]At a seminar in which I spoke on this topic, I was surprised to have an example of this aspect of a utility function--that is, increasing immunity--supplied to me by a colleague. Having had his car stolen (and returned) ten times in the past year, he said he found that the disutility of each additional incident declined as the number continued to increase!

[4]One notable exception--in addition to that of mean-variance portfolio analysis--is the labor-leisure choice of an individual. However, a perusal of that literature in microeconomic theory revealed no discussion of the sign of the second derivative, nor even a convincing argument as to why the indifference curves should bear the shape which is typically assigned them in textbook drawings, i.e. positive first and second derivatives.

[5]This can be verified by substitution of $v^* = \bar{v} - v$ into the earlier assumption, where $\bar{v}$ represents the maximum level of v possible.

[6]Functions satisfying the conditions for first and second derivatives of utility with respect to c and v* may exhibit a number of possibly awkward characteristics in their indifference curves. For example, the sign of their second derivative may be positive for low levels of U, then zero, then negative; it is even possible for the function to have points of inflection along a single indifference curve. For this reason, the utility function considered here is restricted to be one having strictly convex or concave indifference maps. For other cases the analysis can be extended and suitably modified to obtain the desired results, though the task of doing so may be considerably more laborious.

[7]The defendants are thus the "plungers" which Tobin (1958) discussed and then declared inoperative. For those earlier plungers,

however, the choice was between return and risk; consequently, a plunger's attitude toward the bad good revealed quite explicitly his attitude toward risk, namely, that though he disliked it, he became more immune to additional doses as his "stock" of it increased. In contrast, the defendant to which this utility function belongs, is more immune to additional v as his amount in hand increases; and because we have assumed a positive second derivative of utility with respect to v, the defendant is <u>not</u> risk-averse, but rather favors risk with respect to v.

[8]One interesting paragraph in the American Bar Foundation study of several court systems, including the Recorder's Court of Detroit, noted that, "The most common criterion used in selecting attorneys for the indigent is need...'Young and old attorneys needing money will do a very thorough job of investigating, preparing, and arguing the case.'" (McIntyre 1967: 128).

[9]Conceivably, most any of the factors could be related to others pertaining to the case, for example, the seriousness of the offense to the costs of trial; or income with the type and seriousness of the offense, since certain income groups have higher likelihood of being involved in particular classes of offenses (e.g. white-collar workers with embezzlement), some are more likely to be out of jail pending trial, etc. Because introduction of these interrelationships would vastly complicate a model already suffering from that malady, they are omitted here. They could be introduced later, and while they would render the derivation of the results of the model more laborious, their introduction should not substantially alter the results.

[10]Similarly, if an employed defendant goes to jail and his family subsequently goes onto assistance, his disposable resources might actually increase. With the exception of this last possibility, the instituting of release-on-recognizance programs, in which defendants do not have to post bond (but do have to promise to return for trial!) could be expected to increase disposable resources of defendants.

[11]For example, suppose D has no time-dependent income and no legal fees (and that there is no prohibition of his expenditure on case-improvement.) The resource and evidence effects of delay may cause the specialization to reverse under different values of t_v, e.g. f could rotate outward sufficiently to reach $(v_1^*,0)$ as its intercept, and thus result in specialization in v*:

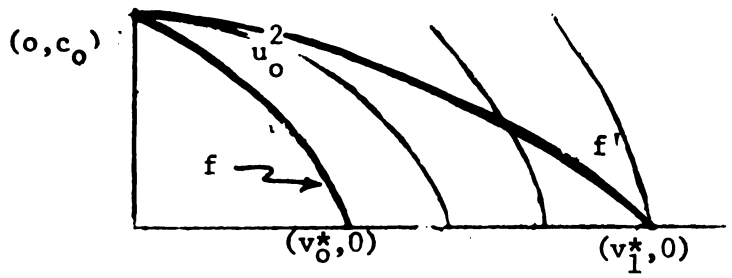


Figure 23. Specialization in v* and possible reversal with time passage.

[12]The derivation of r''_s is an follows: Since we have assumed the signs and magnitudes of the second derivatives to be such that only corner solutions are possible for utility maximization, and since we have fixed the c-intercept of f (by assuming a given level of R) we can determine the level of utility obtainable by specialization in c, which subsequently determines the corresponding indifference curve for that utility level. On that indifference curve, we determine the v*-intercept; and with that information, one simply "solves" the function f for the value of r" for which f passes through the desired v*-intercept.

[13]Normally, the offense to which he would enter a guilty plea would be one having severity less than or equal to that of the original charge. However, under some circumstances a defendant might even plead to a higher charge, although that would be the exceptional situation, if he wished to avoid a label such as that of a sex offender and is willing to take a chance on receiving a harsher (or equally harsh) sentence in order to do so.

[14]Were that not the case, i.e. if $u'_o = u'^2_o$, then the plea constraint would have to lie wholly below the trial constraint, contrary to our understanding of the relation of c_o to c'_o.

[15]This may be interpreted as representing the choice of the defendant who pleads guilty purely to save money, since the outcome is not necessarily better (or worse) than that which he expects from trial.

[16]Such an action is possible and rational for a prosecutor to undertake, for even after conceding to D the outcome D would otherwise expect from trial, P would have conserved his resources and--as is later shown--may still be obtaining a better outcome than what he (P) would have expected from trial.

[17]This is consistent with the idea of there being certain common points of settlement. For example, one of the prosecutors interviewed in Detroit said that in pretrial negotiation, they try not to go below the corresponding attempt charge (e.g. for larceny in a building, not below attempted larceny in a building) unless there are special circumstances which warrant additional concessions. Restated in terms of gains to the defendant, D can expect a reduction of approximately the difference between the original charge and the corresponding attempt charge, with the amount of the concessions being greater or less, depending upon special circumstances and the bargaining skill of the attorneys (or whomever else might be involved in the bargaining.) If the outcome D expected from trial was equal to that obtained from a plea to attempt, he will surely be interested in discussing further the possibility of a guilty plea.

[18]For circumstances which align with those of this third class, the procedure for determining r''_m is as follows: knowing the form and c-intercept of f, one solves for r'' the equation(s) representing g, with the values $v^* = v^*_o$ and $c = 0$ (and possibly also $v^* = 0$ and $c_o = c'_o$, depending upon the form of g) substituted in. The resulting value of r'' in r''_m.

[19]It should be noted that the set of r'' for any of these cases could conceivably be the null set. In particular, if c_o is sufficiently close to c'_o, there will be no r'' satisfying the conditions of case 2. For example, if $c_o < c''_o$ due to a longer pretrial queue, case 2 would revert to Case 1.

2.4 The Pareto Optimal Region and the Outcome of Negotiations

In the preceding two sections, functions were derived which express the minimally acceptable set of bargains for each of the parties considering the option of a plea of guilty. Each function can be seen as dividing the relative strength-outcome plane into two regions, one of which constitutes the set of acceptable plea bargains (for given values of relative strength) and the other of which represents plea outcomes sufficiently undesirable to make trial the preferred method of disposing of the case.

With appropriate adjustment of the variables r' and r" used above to represent relative strength so that each can be stated in terms of one variable r,[1] we can portray these two functions on the same graph and easily view the four possible permutations of acceptability of outcome: (1) Both parties prefer a plea bargain to trial; (2) both parties prefer a trial to a plea of guilty; (3) the prosecutor prefers a plea of guilty to trial, but the defendant does not; (4) the defendant prefers a plea of guilty to trial, but the prosecutor does not. The four regions are indicated in the graph which follows:

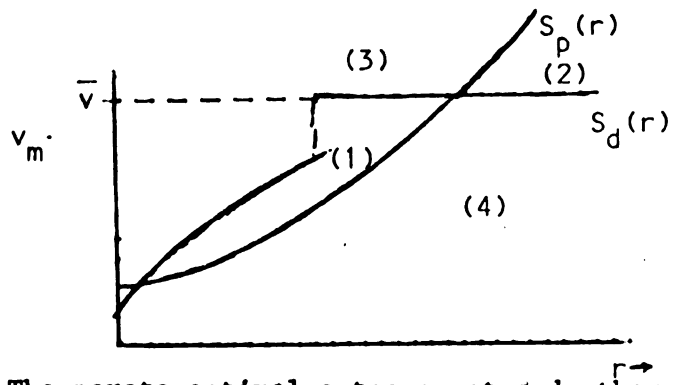


Figure 24. The pareto optimal outcome set and other regions.

Clearly, the pareto optimal region is region (1) and we would expect in general that outcomes would fall within this region. The functions bounding the region are the switch function $S_d(r)$ and $S_p(r)$ derived in 2.2 and 2.3, with both stated in terms of one relative strength variable r.

In addition, in the jargon of section 2.1, the switch functions represent the incremental utility functions at levels of zero increment for the party whose switch function is under consideration. That is, since the switch functions represent the ordered pairs (r,v_m) for which the appropriate party is indifferent between resolution by plea or by trial, the additional utility gained by a plea at that level is zero. Forays into the pareto optimal region from the boundary will yield increases in utility to one party and decreases to the other, but both <u>incremental</u> utility functions will have associated with them positive (or on the boundary, zero) utilities at any point in (1).

For the prosecutor, the incremental utility function is

$$U^*(v_m) = U(v_m,x_m) - U(v_o,x_o)$$

where the notation is as before, i.e. v_o and x_o are the trial outcomes obtainable (and are given dependent upon a certain fixed value of r) and x_m is the outcome on the remainder of P's caseload. For the defendant, the incremental utility function is

$$u^*(v_m) = u(v_m,c_o') - u(v_o',0)$$

and c_o' is the amount of good C obtainable under conditions of a plea of guilty and v_o' is the level of v obtainable via trial. In the event that $u_o^1 > u_o^2$, the latter term on the right would be $u(c_o,\bar{v})$.

The pareto optimal region can then be stated in two alternative ways: It is the set of pairs (r,v_m) such that $S_p(r) \leq v_m \leq S_d(r)$ if $S_p(r) \leq S_d(r)$; or equivalently, it is the set of (r,v_m) for which , for each value of r, $U^*(v_m) \geq 0$ and $u^*(v_m) \geq 0$.

Adopting a split-the-difference theory of bargain outcomes for the reasons cited in 2.1, we anticipate the set of possible bargain outcomes for differing levels of r to be

$$v_m = \lambda S_p(r) + (1-\lambda)S_d(r)$$

where λ:1-λ is the proportional split of the pareto optimal set. That is, we take the outcome function to be a linear combination of the two switch functions, with weights according to the split of the amount of v to be distributed. Further, since the defendant's function cannot be assumed continuous in the general case, but rather is piecewise continuous in the two sectors of its domain, the outcome function needs to be stated in a manner similar to that of D's switch function:

$$v_m = \begin{cases} \lambda S_p(r) + (1-\lambda)\ S_{d1}(r) & \text{for } r'' \leq r''_s \\ \lambda S_p(r) + (1-\lambda)\ S_{d2}(r) & \text{for } r'' > r''_s \end{cases}$$

where S_{d1} and S_{d2} are the pieces of the defendant's switch functions in the left- and right-hand sectors of the domain of r and r''_s is the special level of r" at which D is indifferent between specializing in trial outcome improvement or in consumption.

Of interest at this point in the development of the model is the question of how the Pareto optimal region and the outcome function shift in response to the suggestion of alternative values of the key parameters.

Factors which contribute to the determination of relative strength for a given defendant will determine the actual value of r which is relevant (and consequently determine the specific bounds on the plea possibilities): alternative values of component variables of r cause "movement" along the switch functions and the outcome function. For example, if a defendant adds another arrest to his record during the period pending resolution of the case which we are considering, we can expect this fact to be reflected in a higher value of r, dictating a new reference position on the outcome function. The same would be true for other determinants of r, such as the amount of evidence initially at hand. For these parameter changes, neither the location of the Pareto optimal region nor of the outcome function has changed--only the relevant location in the region or on the function has changed.

However, when two defendants, identical but for the parameter of one of the contributors to r, are compared to one another, the Pareto optimal region and corresponding outcome function would appear to shift. For example, if we believe it more advantageous to be a white defendant than to be a black one, we would expect the region and outcome function for the latter to lie to the left (northwest) of that for the former, as illustrated below.

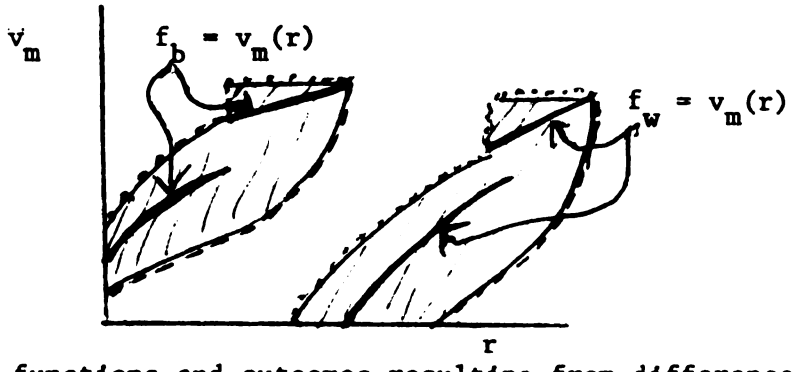


Figure 25. Switch functions and outcomes resulting from differences among defendants. Dotted curves denote D's switch function; dashed curves, P's.

Thus for each level of r, the black defendant will fare worse than the white one, although the outcome will clearly be sensitive to the amount of evidence and the other factors which determine the level of r. To the extent that discrimination exists, it does not supplant consideration of evidentiary and other factors (such as previous record) relevant to court decisionmaking; rather, it shifts the scope of the alternatives to which those considerations are addressed. A similar analysis can be applied to the impact of other characteristics which vary among defendants, and the earlier analysis to alternative situations for one defendant.

Changes in variables introduced through the opportunity set can be traced through the determination of the switch functions to assess their effect on outcome. Other things being equal, supplying the prosecutor with greater resources would translate both the trial and plea opportunity sets upward; not only is the outcome from trial expected to be higher, but the equally-desirable plea bargain increases as well. This is true for each value of r and so the increase in resources has shifted the entire switch function for the prosecutor upward (to the left). Since no such corresponding change has occurred for the defendant, his switch function remains in its former position, and the pareto optimal region contracts. If the proportion of the split remains the same as assumed before, the net effect of the higher level of resources is to shift the outcome function to the left--although not by as much as the shift in P's switch function:

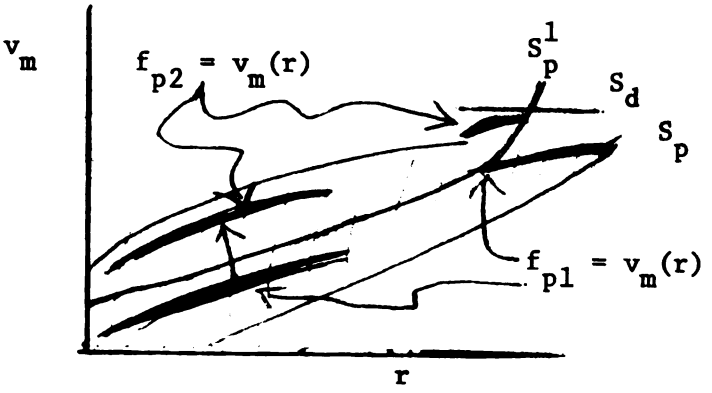


Figure 26. Leftward shift in outcome function due to greater prosecution resources.

Similarly, increases in defendant disposable resources, through negative tax or subsidy programs or whatever, would shift S_d to the right, resulting in an expanded bargain region and an outcome function to the right of its original position.

As noted in the section on the defendant, the effect of time cannot be specifically determined without quite a bit of information on the defendant, and then could vary among defendants. The effect on the prosecution is to shift his switch function to the left: neither the cost effects nor the evidence effects of delay is helpful to him. The defendant, on the other hand, may find one or both of those effects to be beneficial; we especially expect the evidence effect to favor him. Suppose that the effect on the defense is damaging: S_d shifts to the left. Since S_p has shifted to the right, the combined effect from both parties is difficult to determine and will depend upon the comparative damage inflicted by the delay. On the other hand, if delay is helpful to D, his switch function shifts to the right along with P's and the

outcome function does so also. This latter situation seems to be the belief among law enforcement officials as well as among defense attorneys.

Continuing the same manner as above, it is possible to analyze the effect of differing circumstances confronted by the prosecution and the defense. Those which favor the defense shift his switch function (and ceteris paribus, the outcome function) to the right; ones which favor P shift his switch function (and ceteris paribus, the outcome function) to the left. Depending upon the joint effect on the two switch functions, the Pareto optimal region itself shifts, expands or contracts, or does both. With the implicit assumptions involved in adoption of a split-the-difference theory of outcome--especially those concerning the probability distribution of outcomes within a bargain interval for a fixed value of r--one can posit the outcome function to be descriptive of the set of possible settlement points for pleas of guilty.

NOTES to 2.4

[1]Further discussion of possible transformations follows later in this section and the next chapter. For the moment, let one example suffice: if we define $r = -r' = -1/r''$, we can speak of the outcome as viewed by each party in terms of the variable r alone.

3. A Model of Plea Bargaing: Empirical Testing and Results

3.1 The Transition to an Empirically Testable Model

In the preceding chapter a model was posited to determine the bounds on the outcome of mutually agreed upon pleas of guilty. In addition, under the assumption that the amount of v within the pareto optimal interval was to be split between the parties in a fixed proportion, an outcome function for v in terms of r, the relative strength variable, was determined. In this chapter both of these descendants of the switch functions are subjected to empirical testing using the specific relations introduced in earlier sections as examples for the prosecution, the defense, and the outcome of two-party negotiations.

The first task, that of testing the validity of the switch functions as bounds on the set of plea bargain outcomes, presents significant problems due to the lack of appropriate statistical tools. Much of the thrust of applied statistics and econometrics has been devoted to the search for ways of expressing relationships in terms of measures of central tendency and corresponding measures of dispersion from them. In contrast, little attention has been paid to the complementary issue of placing outside bounds on sets, and attempting to characterize relationships by functions which (together) contain the set of relevant points, rather than by one function which approximates them.

There are two notable exceptions to the indictment of the paragraph above. The techniques of quality control and its subsidiary concerns attempt to define bounaries on regions of acceptable outcomes, but require foreknowledge of the function whose approximation the

bounding functions enclose and the maximum acceptable tolerance. Hence these are really problems of finding boundaries subject to constraints which are tied to previously known functions describing (desired) central tendency: the bounding functions are determined from information (or requirements) on central tendency, rather than from the data directly. The second possible resource for determining bounding functions is a technique suggested by Michael Farrell, who determined an isoproduct curve by connecting the set of outside points (i.e. ones lying farthest to the southwest on a graph), using crossectional averaged data on manufacturing. [1] However, his method clearly could be disastrous for use with rougher micro cross-sectional data, in which one or more outliers could unfavorably affect the results--as would be the case if applied to data on the trial versus guilty plea choices made by individuals. Thus, the statistical literature does not provide much guidance for determining the validity of the theory of the functions which bound bargained outcomes. Nevertheless, some graphical evidence is presented in which cases are segregated by different criteria and the relative strength-plea outcome relationship can be observed.

The second task, that of testing the formulation of the outcome function derived from a split-the-difference theory of negotiation, is a bit easier and is performed using multiple regression, following the imposition of some additional simplifying assumptions. Since the function to be estimated is in general expected to be discontinuous--at least nondifferentiable at one point--a modified search technique is employed which is similar to the approach of Quandt (1958, 1960) in

his work on shifting regression regimes for use in determining the point of discontinuity of the outcome function.

As developed in the second chapter, the outcome function for a split of the Pareto optimal intervals in proportions λ: $(1-\lambda)$ was determined to be

$$v_m = \lambda S_p(r) + (1-\lambda)\, S_d(r).$$

Since S_d is defined differently when D is specializing in trial outcome improvement than when he is specializing in consumption, the outcome function must be stated differently for these two circumstances also:

$$v_m = \begin{cases} \lambda S_p(r) + (1-\lambda)\, S_{d1}(r) & \text{for } r \leq r_{disc.} \\ \lambda S_p(r) + (1-\lambda)\, S_{d2}(r) & \text{for } r > r_{disc.} \end{cases}$$

where $r_{disc.}$ is the level of r corresponding to min $\left\{r''_l, r''_s\right\}$ and S_{d1} is the defendant's switch function to the left (i.e. for low values of r) of the discontinuity, and S_{d2} is his switch function to the right of that point.

For the particular utility functions and opportunity sets posited in earlier sections, the outcome function is

$$v_m = \lambda \frac{(R/w)^{(\alpha+\beta)/}}{(R'/w)^{\beta/\alpha}} \left(\frac{\alpha}{\alpha+\beta}\right) h_p(r) + (1-\lambda)\, \bar{v}$$

$$+ (1-\lambda)\, \gamma\, \underline{R}\, h_d(r)$$

where $\gamma = 1$ for $r'' \leq r''_{disc.}$ and $\gamma = 0$ otherwise; where $h_p(r)$ and $h_d(r)$ are defined so that h_p increases and h_d decreases with increases in r; and where the definitions of all other terms are as before. As derived in the example of section 2.3 the point of the discontinuity of S_d and

hence of the outcome function is

$$r''_{disc} = \frac{[(\bar{v}+1)\ \ln(c'_o+1) + 1]}{(\bar{v}+1)^2\ \ln(c'_o+1)} \cdot c_o \ .$$

Since the function as stated above contains too many parameters to be estimated at once, it is necessary to make assumptions concerning the values of some of them, and about functional relationships among others. Once those assumptions have been made, the outcome v_m is a function of γ, R, R', $\bar{v}$, $\underline{R}$, and r, all of which have known values for a given defendant, subject to the adoption of some additional assumptions such as those involved in sorting out variable from fixed costs in the prosecutor's budget.

With the exception of γ, each of the variables appearing in the outcome function is subject to scaling problems. For the two representations of the prosecutor's resources, the problems (to the extent that they exist) may be traced to assumptions made in determining the levels at fixed, attributable, and variable costs, as well as to the definition of w, the price per unit of the good introduced as variable input to evidence production. Although it would be desirable to estimate the values of α and β, rather than to assume them, it is not possible to do that with the information available, so the assumptions entailed there are another source of possible error.

Similarly, judgements about the disposable resources of the defendant under conditions of a trial or a plea bargain pose scaling problems due to the initial incompatibility of most income data with the definitions of $\underline{R}$ and $\underline{R}'$ employed here. The adaption for use in applying this model requires assumptions about its utility function as

well as about the time period relevant to his decisions. The scaling problems surrounding the use of $\bar{v}$ are less obvious: recall that $\bar{v}$ was defined to be the worst outcome (as viewed by D) that could occur at trial, defined by his trial constraint. In using data on maximum or average sentences applicable to all defendants, we may not be using the appropriate value of $\bar{v}$ for a particular defendant, or for the particular charge against him. [2]

Finally, the introduction of the h-functions poses problems not only of scaling but of multicollinearity among the terms. Since no theory of the determinants of relative strength has been proposed here--its contributing variables have been listed but not integrated within an economic theory--it is impossible to say what a correct measure of that variable should be. The familiar question of ordinal versus cardinal measures is also relevant here. Nevertheless, an index of relative strength is utilized in the estimation; the construction of that variable is discussed in section 3.2.

The spectre of multicollinearity appears because the same source of information about relative strength is used in both the first and third terms. This is in contrast to the theory of 2.2 and 2.3, in which the first term uses the defendant's subjective estimate, and the third term uses the prosecutor's. However, the use of one measure is necessary not only for obtaining the required data, but also within the theory (as mentioned in 2.4) for permitting the superimposition of both switch functions upon the same space. The multicollinearity exacts its toll in the variance of the estimates of coefficients, which may or may not be a hardship.

However, the appearance of r in both terms also adds some complexity to the determination of $\partial v_m/\partial r$, the incremental effect upon outcome of a stronger prosecution case. This is discussed further in section 3.3. Let it suffice to say here that with the exception of $\partial v_m/\partial \bar{v}$, none of the partial derivatives is a constant--and thus able to be read from the estimated equation as the coefficient of a lone term in the regression--but rather, each is itself a function. Consequently, determination of the signs of partials in order to verify from empirical results the contentions of the theory must involve additional computation. Nevertheless, if those computations are performed, some inferences can be made about the effects of a number of variables upon plea bargain outcomes.

Also, momentarily ignoring the possible scaling problem with $\bar{v}$, it is possible to read directly from its coefficient the proportion of the bargainable difference accruing to each of the parties. Since the coefficient λ is the proportion of D's reservation price which P is able to exact, values between .5 and unity would suggest that the prosecutor is faring better than is the defendant.

The results of the above tests lend themselves to interpretation both in terms of the validity of the model as a whole, and in terms of the value of the parameter λ which describes the proportion by which the two parties split the difference between their reservation prices.[3]

The data available, problems encountered in the empirical work, presentation of the results, and their interpretation appear in the following sections.

NOTES TO 3.1

[1] Farrell chose the points to connect with one another by checking to see whether inclusion of the i-th point would cause the new set of connected points to jut in and out again, i.e. he required that for the i-th point to be on the boundary, the function thus obtained must retain its convexity with respect to the origin. Since he used relatively nice averaged data, his results presented a semblance of reasonableness, thus recommending the technique for use in similar settings.

[2] For example, if it is customary for any offense carrying ten years or less for a plea charge to carry half or less of the penalty on the original, then for those offenses, the appropriate value of $\underline{v}$ would be one half of the maximum on the original charge. This sort of problem cannot be avoided where indeterminate sentences are used, without going to the (possibly inaccurate) extreme of interviewing defendants to determine their opinions about $\underline{v}$.

[3] The results on the latter question provide an interesting comparison with findings of Daniel Hamermesh (1973) in his test of a split-the-difference model applied to labor-management contract negotiations. Among other problems associated with the methodology used, Hamermesh tested only for a split of exactly fifty-fifty. Nevertheless, he was able to reject that hypothesis, finding that final settlements were much closer to management's initial offer than to labor's. It would be helpful to test his theory in a manner which would search for the proper proportion to describe the split, or repeat the procedure he used under alternative proportional splits. Since this is the only empirical work of which I am aware that uses a split-the-difference theory, there are no others to which the results determined in the present work could be compared.

3.2 The Data and Its Discontents

Because much of urban crime--crimes directly affecting the physical state of persons and property (as opposed to more exotic offenses such as skyjacking, or crossing state lines with intent to incite a riot, which are federal offenses)--is the purview of state and local law, usually administered by county or district courts, respectively, there exists the potential for great variation from one institutional context to another. Consequently, the transition from the theoretical model to one to be tested in a specific institutional setting must take into account the peculiarities of applicable laws, rules, and traditions. [1] This section includes a brief sketch of the Recorder's Court of Detroit, from which the data used in the following section was obtained; it also contains a description of how the data were collected and utilized, a summary of some aspects of the individuals whose cases were drawn in the sample, and finally, a list of other steps taken to prepare the data for use in testing the theory of plea bargaining. [2]

Recorder's Court, born in nineteenth century America, is beset in the twentieth century by a spectrum of challenges and problems not unlike those found in other major cities. In fact, one might even say that the Detroit Recorder's Court is close to the prototype of the urban court. [3] Its geographic jurisdiction encompasses the portion of Wayne County which is also within the city of Detroit (hence excluding most of the suburban and rural areas nearby); as one of two county courts (the other in the nonDetroit portion of the county), its

case jurisdiction is over offenses against the laws of the State of Michigan. Along with many other American courts, Recorder's Court found itself in the mid-sixties mired with an excessive backlog of cases, facing complaints from all parties to the operation of the criminal justice system, including the public, police, defendants, prosecutors, lawyers, and others. The response to this dissatisfaction was almost as diverse as the problems which spawned it, as visiting judges were brought in to hack away at the backlog--outnumbering the regular judges by three to one--and much general streamlining of procedures was begun which is at least acceptable, if not commonplace, today. However, this reorganization has not eliminated the collection of problems which initially arrive on the court's doorstep, nor changed the basic processes which are carried out by the prosecutor and the court.

One major life process of a court is the practice of plea bargaining, the method by which the great majority of a court's cases are disposed. For most of its history--which in the United States dates back to the Salem witch trials[4]--plea bargaining has been a process confirmed in law only by its mere existence, rather than by any justification in statute law or case precedent. Although the last three years have seen a change in this perspective on plea bargaining as an extra-legal process, [5] that metamorphosis was only beginning at the time Recorder's Court initiated its reforms of the plea negotiation process.

Although each case receives initial processing by several of a large number of persons involved in entry-level research and evaluation of cases, the actual discussion and possible settlement of cases before trial is handled by a small unit of three persons who together with

supportive staff, specialize in exactly that work. Every felony case is scheduled for a pretrial conference, as the session is called, and although a defendant may request that no reduced plea be entered (if he does not wish to consider a plea of guilty) most defendants or their counsel attend the meeting, and most subsequently plead guilty. The prosecutors who work in the Pre-Trial Division, as well as other participants in the criminal justice system, believe that the "new" (seven-year-old) system breeds greater uniformity in the handling of pretrial settlements, as well as greater opportunity for flexibility when warranted by the circumstances of individual cases. This flexibility ranges from the very deliberate consideration of factors relevant to the rehabilitative potential of the defendant, such as his age, previous criminal record, relationship to the labor force, etcetera,to the special difficulties experienced by out-of-towners in a city far from home.

However, the three pretrial prosecutors also operate under a set of rules concerning the general handling of cases in the absence of special circumstances. Those rules, together with the coordination possible in such a small group in which one member also works as an administrator of the other two,provide for greater accountability than was possible under the earlier setup in which a hundred prosecutors carried out those tasks.

In addition to introducing centralization of plea negotiations, the advent of the pretrial conference as an integral part of the trial process was accompanied by other changes as well. For those defendants who decide to plead guilty, the case can be sent to (abbreviated) trial

in a much shorter period than would otherwise have been necessary. In fact, for the period relating to the data used here, it was the practice of two judges to handle such cases more or less full-time, [6] thus creating a somewhat separate set of machinery for the remainder of the court's processing of the guilty plea-defendant's case. The alternative routes for a case are sketched in the possibility tree diagram below, along with the percentages of cases along each branch. [7] Figures in parentheses denote percentages of those considered at the point of branching, that is, they refer to the total caseload.

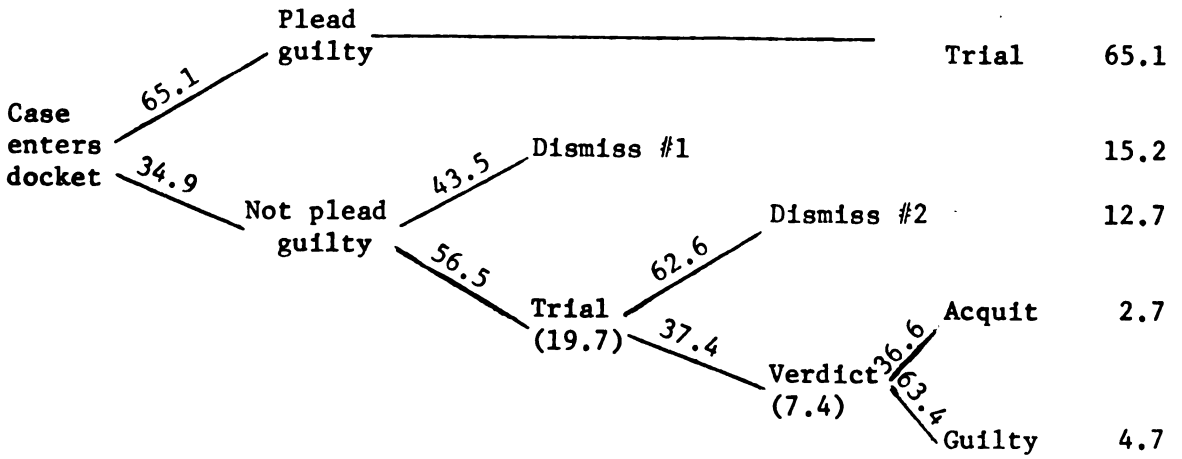


Figure 27. Alternative paths for a case and distribution of cases by mode of resolution.

Dismiss #1 denotes methods of dismissal before trial, which are dominated by dismissals upon motion of the prosecutor. Trial in the case of the defendant pleading guilty is a short and usually perfunctory affair, and should not be equated with trial in the case of the defendant who does not plead guilty. Dismiss #2 relates to dismissals after

cases have entered the trial process--having left the sphere of prosecutor discretion--and are derived principally from cases in which dismissal was made by the judge (not upon recommendation of P). Verdict represents the situation in which a guilty or not guilty verdict is rendered, and the last two branches are self-explanatory. It should be noted that, although the Annual Report does not separate the subgroups, defendants may be found guilty of the original charge or of any lesser included ones, [8] so that the clumping of these together implies some loss of information concerning the alternative which a defendant confronts when considering trial as an option. It is also worth mentioning that if there is any likelihood that guilty defendants tend to plead guilty more often than do not guilty ones, the lower paths in the preceding sketch will represent a more rosy picture than an arbitrary defendant would face.

In Recorder's Court, the object of bargaining, exchanged for the promise of a guilty plea, is usually a reduction in the charge to which the defendant would plead; the average reduction for defendants in the data set used was to slightly less than four-tenths of the original charge, where the ratio taken is that of the maximum penalties of the plea charge to original charge. However, concessions may take other forms as well, forms which represent "in kind" transactions and which did not appear in the case files drawn in the sample. For example, it is not uncommon for sizable concessions to be made by a prosecutor upon assurance that a young defendant is willing to enter military service,

in order that a prison sentence not prevent or deter that activity.[9] Because it is possible for the entirety of concessions by either side not to be fully apparent from perusal of case files, the results derived from such data should be accepted with a small bit of reservation; however, they do tell most of the story.

The information used in this study was drawn from the docket sheets of the Pre-Trial Office, and corresponding files of the Prosecutor (the main files). The first forty property cases were drawn from the pretrial docket sheets for each of February, April, June, and August of 1972 to comprise the original sample. Although details regarding the composition of this group are presented in the appendix, it is noted here that the choice of dates was designed to largely avoid seasonality problems (such as more of certain crimes at Christmastime, etc.); to be after the initiation of a standard evaluation form providing most of the informational items of use here; to be recent enough to be of some relevance for policy consideration; and sufficiently old for most of the cases to be closed, thus providing information on disposition.

It was decided that the sample should be restricted to felony property cases for several reasons. Paramount among those reasons was the heavy reliance necessarily placed upon the complainant in personal injury cases: if the complainant refuses to cooperate at any time, the case is essentially over. Further, for a number of offenses, such as with rape, the complainant must be exonerated before the defendant's guilt can be considered, thus adding a "stage" to the process not included in the original model. Finally, since there is some

speculation as to the rationality of offenders in many classes of personal injury offenses,[10] it was not clear that the model's most basic assumptions could be met for randomly selected personal injury cases.[11]

Although the recordkeeping of the prosecutor's office in Detroit stands out among American cities as quite thorough, certain items of critical interest here were occasionally left blank on the forms; because of these missing pieces of information and because of some other minor problems detailed below,[12] it was necessary to eliminate some observations from use here. For the one hundred fifty-five cases from which the description was drawn, a miniature new year was posited, with the prosecutor's budget scaled accordingly (as best could be estimated[13]) and the prosecutor's utility function was assigned parameters $\alpha = 1/155$ and $\beta = 154/155$ to reflect equal (multiplicative) case weightings. It was further assumed that the costs of trial were composed of the costs of executing the trial itself plus the costs per day to the prosecutor of maintaining the case until the dates of the pretrial conference, and the subsequent costs until trial.[14] Together with information on income derived from the 1970 Census, attributed to defendants based upon residence address zip codes,[15] the information available was sufficient for engaging in empirical testing.

Some characteristics of the defendants in the sample who pleaded guilty--with dismissal considered a guilty plea at outcome zero--were as follows: the average value of v_o, taken to be the maximum penalty[16] on the original charge, was 15.1 years; the average of the maximum penalties on the reduced plea offenses was 6.3 years, with the average ratio of the latter to the former being .36 as alluded to earlier. The distributions of original and reduced charges and the maximum sentences for each are shown in tables one and two.

Table 1. Distribution of defendants by type of charge (in percentages).

	Original Charge	Plea Charge
Felonies		
Auto-related	8.4%	.6%
Robbery	27.7	4.5
Burglary	31.6	8.4
Larceny	32.3	44.5
Misdemeanors		31.0
Dismissals		11.0

Table 2. Distribution of defendants by length of maximum sentence (v) (Distribution is in percentages.)

Years Maximum Sentence[a]	Original Charge	Plea Charge
0 (dismissal)		10.3%
.2[a]		10.3
1		1.3
2	.6%	34.8
2.5	2.6	12.2
4	4.5	6.5
5	36.2	11.0
10	20.6	1.3
15	12.9	1.9
25[b]	1.3	.0
life	21.3	10.3

[a]Ninety days was coded as .2 years

[b]For two defendants, the sum of two separate offenses on same case was 25.

Over half of the defendants, 53.5 percent, awaited their trials in jail, in spite of the release-on-recognizance program to which the court subscribes. Of the defendants in custody, the proportion of blacks [17] was within a half-percent of their proportion in the total sample, which was 77.4 percent. For slightly over a quarter of the defendants a weapon was involved in the associated incident, thus providing--in addition to the distribution of maximum sentences--an indication of the seriousness of the offense.

As a group, the defendants were young, with most falling into the eighteen-to-twenty-four age group: 17 percent were juveniles (under age 18), 76 percent were age 24 or younger, and 90 percent were age 40 or younger. In the great majority of instances--over 88 percent--the complainant and defendant were not related by family ties or by friendship. [18]

Although 24 percent had at least one other case pending in the court, only a quarter of those defendants had two or more. It was the policy of the court if a defendant had other cases pending, that all his cases be handled at once when the first reached the head of the queue. In practice, negotiations then dealt with the entire $\sum_{j=1}^{n} v_{oj}$ where n is the number of cases (including the one appearing in the sample) for a given defendant, rather than simply being concerned with v_{o1} which we observed. Because it was typical practice for the defendant involved in several cases to plead guilty to one offense and have the others recommended for dismissal, the resulting v to which D pleaded could be expected to exhibit greater variability for these cases: either D would plead guilty to another offense not in the sample and

the observed outcome would be zero in spite of the presence of other factors which would tend to make us expect a positive outcome; or the value would be particularly high relative to the original charge, and the observed value of v_o would result in greatly understated value(s) for the dimension(s) of the Pareto optimal set. [19] Hence the effect of other pending cases of an individual defendant is to increase the variability of outcome on the case included in the analysis.

For the group of defendants in the sample, the average time from the date of the offense (which is usually within a day or two of the arrest date) to the pretrial conference was 70 days; however, the median was slightly over 27 days, indicating that most cases reached that point in far less time than the mean would lead one to believe. From the time of the pretrial conference to the trial date available with the (later-) assigned judge was 72 days on average. Because this interval was not necessarily (or even usually) known to defendants before the pretrial conference, [20] the average value was used in the empirical work which follows.

For use in the testing of the model, the sample had to be pared somewhat. The sample included only five women, which was an insufficient number upon which to determine the impact of the sex of the defendant upon the outcome of plea negotiations. Further, since women are reputedly treated with some deference by law enforcement agencies, failure to take account of that factor could be detrimental to the validity of any test of the theory, as well as proving deceptive should one desire to use the model for prediction.

Second, observations were eliminated for which values of critical

variables were missing from the data. The most prominent among these "hatchet" variables was the prosecutor's estimate of the probability of conviction, which is crucial for the determination of the relative strength variable. In addition, for what one might call "two-plus" defendants (where a case had more than one defendant), the value of that piece of information was of questionable reliability, due to the manner in which items were recorded for defendants other than the first one listed for a given case, [21] so "two-plus" defendants were also excluded. Had those cases been maintained in the sample, they would have imposed a curious weighting on the types of cases considered. (For example, eight persons were accused of breaking and entering a railroad car while it was still in the station, an otherwise unusual occurrence.) Finally, observations for which the disposition of the case was not given or for which it was sent to trial (and hence could not have been resolved by pretrial settlement) were deleted. Dismissals were kept in the sample, and assigned the value zero for outcome. The final set of observations used consisted of eighty-six "first" defendants.

It was extremely fortunate for the empirical testing of the theory that the Detroit data included an estimated probability of conviction, for had there been no such information available concerning the priority or strength of the case as perceived by one party or another, it would have been necessary--and perhaps, quite precarious--to construct such a variable, and to do so independent of any concept of how either party makes such evaluations. Since information on the prosecutor's assessment of the probability of conviction was available, it

was possible to derive from this a function which could be used as an index of the relative strength of the two parties. The derived index of relative strength $\hat{r}$ (derived below) was believed preferable to the prosecutors' probability estimates for use in later regressions for several reasons. It seemed to be the pattern of the prosecutors that estimates were always made in multiples of five, i.e. 60, 65, ...,100 [22] (with the exception of one modest prosecutor who wrote 100's as 99) and that pattern made it difficult for such a proxy for r to be rightly considered a continuous variable, a requirement inferred by the theory. Since greater specificity on the part of the prosecutors would likely have provided only spurious accuracy anyway, the scaling system they used probably served its administrative purpose well. However, for our use of r as a continuous variable and in order to decrease the harsh lines otherwise drawn between cases in which the "true" r-values (as perceived by a given party) are close to one another, it was decided to use the predicted values of the probability of conviction, stated as a function of the variables mentioned earlier as being relevant to the determination of relative strength: jail status, race, race-times-jail, total time from offense to trial, an index of the harshness of the judge to which the case was assigned for trial, [23] whether the defendant had other cases pending, whether the defendant and complainant had family or friendship ties, and whether the defendant was a juvenile. The resulting predicted values for the probability estimates were used as an index of relative strength, and therefore should not be interpreted as simply representing probabilities.

Nevertheless, in order to retain the option of interpreting r in

such a way as to have the values bounded--which the probabilistic interpretation conveniently required--it was necessary to use a bit more elaborate function than one which was linear in each of the variables listed in the preceding paragraph. A logistic function was used, having the form

$$r_p = \frac{1}{1 + e^{q(z_1, \ldots, z_d,)}}$$

where the z_i are the characteristics of the defendant and of the offense, r_p is the probability variable from the prosecutor's files, and the function $q(z_1, \ldots, z_d)$ is assumed to be linear in the parameters.[24] The numerator is the upper bound on the function, which we have taken here to be one.

By suitable algebraic manipulation, the logistic function can be stated so that the exponent of e forms the right-hand side of the equation, and the left-hand side is $\ln(\frac{1-r_p}{r_p})$ which is the natural logarithm of the odds in favor of the defendant. This transformation of the logistic function permits estimation of a linear function rather than of a nonlinear one. And since the logarithm is an increasing function, the results of the regression in which the dependent variable is the log of the odds in favor of D lend themselves to easy interpretation in terms of the relative strength variable, for the ordering is exactly opposite to that favoring the prosecutor. Once the coefficients for the z_i have been determined, it is a simple task to transform them back to the form of the logistic function and use the estimate of r thus obtained in any subsequent work.

The results of the regressions on $\ln(\frac{1-r_p}{r_p})$ are presented below,

with alternative specifications of the q-function to be found in the appendix (the alternatives arising from different representations of the age variable, as well as inclusion of some alternative variables.)

$$\ln\left(\frac{1-r_p}{r_p}\right) = \underset{}{-1.13} \underset{(-3.55)}{-1.24\ \text{Jail}} \underset{(-1.98)}{-.55\ \text{Race}} \underset{(2.63)}{+1.05\ \text{Race Jail}} + \underset{(1.07)}{.002\ \text{Total time}}$$

$$+ \underset{(.56)}{.004\ \text{J Plea}} \underset{(-1.94)}{- .36\ \text{Pend Chg}} + \underset{(1.16)}{.30\ \text{Know Wit}} \underset{(-.98)}{- .22\ \text{Age 17}} + e_i$$

$$R^2 = .23^{225}$$

where the variables were given values as follows:

Jail equals 1 if D is in custody and 0 otherwise:

Race equals 1 if D is nonwhite and 0 otherwise;

Race Jail equals 1 if D is nonwhite and in jail, and 0 otherwise;

Total time is the number of days between the offense and pretrial conference, plus 72 (the average from pretrial date to trial);

J plea indexes the harshness of the judge by the percentage of defendants assigned to her or him who plead guilty (and so J Plea decreases, the harsher is the judge);

Pend Chg is 1 if D has other cases pending and 0 otherwise;

Know Wit is 1 if D and the complainant are friends or relatives and 0 otherwise;

Age 17 is 1 if the defendant is age 17 or younger and 0 otherwise;

and t values are given in parentheses.

Although not all the coefficients are significant at low significance levels, most signs are as expected. Jail status and race have the anticipated signs, with the coefficient of their interaction term indicating that (1) the use of only the two variables (i.e. excluding the interaction term) would overstate the effect on relative strength; and

(2) including the joint effect of race and jail results in the following order with respect to the defendant's strength vis-a-vis the prosecutor, in order from strongest to weakest: whites not in jail, blacks not in jail, blacks in jail, whites in jail. The order of the latter pair of sets has several possible explanations. One is that anti-black discrimination in pretrial incarceration results in having blacks with strong cases residing in jail while their white counterparts are released. Another possible explanation is that of discrimination in favor of blacks by the black-dominated juries in Detroit; this would be consistent with some previous speculation about pro-black bias of juries (Time, April 2, 1974) but would seem to conflict with the ordering of persons not in jail, unless the discrimination is based upon other factors as well, such as income, social class, or age. Finally, the functional forms chosen and data used may be a source of this anomaly; however, the significance of the coefficient possibly makes this caveat less necessary.

Delay is beneficial to the defendant, as is having a less harsh judge; in contrast, having other pending charges or being age 18 or older would work to D's disadvantage. The sign of Know Wit would seemingly be negative, since positive identification by a relative or friend would appear to be an obstacle to D's defense; however, since the estimate is being made by the prosecutor (recall that all of this is based upon his estimate of the probability of conviction) it is possible that he would consider the identification by a disinterested party -- who would be less likely to recant at the time of trial, having little sympathy for D -- to be more dependable.

The importance of the jail status of the defendant is deserving of special note, for it recurs elsewhere as well as here. Its impact on the prosecutor's estimate of conviction, and consequently upon the estimated relative strength variable, occurs for several reasons. The first to come to mind is likely to be the resource effect; for a defendant in jail is less able to earn, borrow, or otherwise obtain resources for use on his defense; and if determined to be indigent (which may result in part by his being in custody) and assigned counsel, he may be considered worse off than were he able to choose his own attorney.[26] However, the resource effect is accompanied by other factors having perhaps even greater impact: (1) The defendant who is out of jail is able to work as a surrogate employee of his attorney, doing some of the legwork of preparation for trial which might otherwise have been ommitted, or accomplished by paid employees at a higher cost to D. (2) The defendant is able to perform a number of tasks which would be difficult if not impossible for others to carry out, such as recognizing witnesses by face (a particularly valuable ability if their names are not known) and locating them in places they frequent, for D is more likely to know their friends and habits than would his attorney. (3) Finally, the effect upon juries at trials in which a defendant is led in by a bailiff, rather than arriving of his own volition, is to confirm the attitude that "he wouldn't be here if he weren't guilty," a belief that is surely unfavorable to the defense. Thus the strong and significant negative impact of the defendant's being in jail is not a surprising finding.

With the determination of the index of relative strength, we have sufficient information with which to estimate the outcome function for plea bargains.

NOTES to 3.2

[1]One notable example of the contrast among various institutions is provided by comparison of pretrial negotiations in Michigan with those in Kansas, as pointed out in the American Bar Foundation study of several state and local law enforcement systems. In Michigan--and in Detroit, the source of the data used in this study--the offense upon which arrest was made is carried forward from one stage of the criminal justice process to the next (possibly with additional included charges tacked on) until a pretrial settlement is reached, at which time the prosecutor would take the action necessary in carrying out his part of the agreement. In Kansas, however, the plea bargain is made earlier in the process, and a researcher unaware of that would be surprised to find that all defendants who plead guilty do so at exactly the level of the "original" charge recorded in the prosecutor's files. Clearly, in the latter situation, a data collector would have to delve further back into the process to determine the original level of v to which the level of the guilty plea should be compared. (MacIntyre 1967, Newman 1966).

[2]For more detail on the evolution of these modifications, the reader may wish to consult the appendix containing the more grubby aspects of that process, such as the step-by-step breakdown of the various costs associated with trial; for the preservation of reader interest and sanity, the more laborious calculations are bypassed here in favor of a summary presentation.

[3]In fact, as such, Recorder's Court was chosen for one of two special case study sites for the Task Force on the Courts report to the President's Commission on Law Enforcement and the Administration of Justice. See The Challenge of Crime in a Free Society, by the Commission and the Task Force Reports: The Courts.

[4]One man was crushed to death under stones, while being offered before each additional stone was dumped upon the existing pile, the opportunity to confess his witch status and thereby be released.

[5]Some of the case decisions on this have been cited in the preceding chapter. One of the more important was Santobello v. New York in which the Supreme Court ruled that a bargain made by a prosecutor must be executed by him or his successor. In their discussion of the case, the justices cite the necessity of plea bargaining to the operation of the criminal justice system, particularly because a requirement that every case be tried would result in hopeless backlogs.

[6]This and other similar generalizations in the text come from interviews with two of the pretrial prosecutors in Detroit, from the Annual

Report of the court for 1972, and from reading and study in criminal justice. Where possible, more detailed citations are given.

The reader can verify this statement on the allocation of judges' time by consulting the statistics on judges in the data summary appendix and noting the small percentage of time for a few judges spent in trials (jury cases plus "waivers".)

The system of "plea judges" was changed in September 1972, when a blind draw docket system was instituted. Under the new system, each entering case was assigned to a judge (randomly determined) and placed in her or his trial queue. The change in procedure was in part a response to the problem (for the court) of judge-shopping, in which some defendants would attempt to delay their cases, renege on plea settlements, and utilize other steps in the hope of being assigned to another judge. (Annual Report, 1972)

[7]See the Annual Report for 1972 for the raw figures, which are also listed by finer categories of offenses. The percentages used in the diagram were derived from those figures.

[8]For example, a defendant accused of first degree murder may be found guilty of second degree murder.

[9]Another example of in-kind settlements is provided by the settlement of some less serious offenses charged to out-of-towners: if they wish to plead guilty early on (at arraignment, typically) they may receive suspended sentences on the condition that they promise not to return to the city. (A bit like in the Westerns, as one seminar observer pointed out!)

[10]This caveat is in keeping with the now-established tradition in the economics literature of excluding individuals who are not sure to be rational, as that term is defined by economists. See, for example, "Crime and Punishment: An Economic Approach" by Gary Becker (1968).

[11]Drug cases were also excluded for the same reason; however, since most drug offenses are violations of federal law, only two observations were affected.

[12]A number of the data problems were within repair, as for example, when the question about whether a weapon was involved was left blank but the defendant was charged with armed robbery. However, a number of items which would have been interesting to try as indicators of the amount of evidence for the prosecutor were not always supplied, for example, whether the defendant had made a confession.

[13]The set of cases drawn for this study constituted 1.33 percent of the felony cases in 1972, which corresponded to a budget of $58,551 for non-fixed costs for the miniature year. Totals to which

the part-year percentage was applied were given me by telephone from the Budget Office, County of Wayne.

[14]Based on the assumptions listed in the appendix, the average cost of trial was estimated at $1817 and a per day case maintenance cost of one dollar was assumed, since various alternative intermediate assumptions showed that to be a ballpark estimate and the true figure was impossible to determine from the limited information at hand.

[15]Information on the defendant's income, welfare status, and employment are given on the Pre-sentence Report, to which I did not have access. To substitute for this, the zip codes of defendant's residences were noted and levels of income were assigned to defendants from the 1970 Census based upon zip code. However, the Census does not give a median or even a meaningful mean income by zip code (the mean income figure includes quite heterogeneous units) and so the numbers used were the proportions of the individuals in the zip codes who had incomes above the poverty lines determined by the Social Security Administration. These numbers were then adjusted to represent reasonable dollar figures, additional assumptions were made concerning the amount of savings at hand, and the defendants' disposable resources for trial and guilty plea alternatives were computed using the expressions derived in section 2.3.

[16]This is another example of differences in institutional practice which must be taken into account in the type of empirical analysis performed here. From statements of persons in the prosecutor's office and from the records I used, it was evident that the relevant consideration in Detroit for negotiation of guilty pleas is the maximum penalty associated with an offense, (with the exception of the "labeling" problems associated with sex crimes, child abuse, and a few other offenses.) In contrast, according to one prosecutor, it is the practice in New York and Chicago for plea negotiations to be concerned with the actual sentence to be given, since prosecutors there frequently make recommendations on sentence (or agree not to make a recommendation, as was the arrangement with Mr. Santobello.) Although information on maximum penalties might be available for the latter two cities, it would not be the relevant figure for use as the plea outcome, and probably not for the original charge either.

[17]As is shown later in this chapter, the jail status of a defendant overshadows the racial factor in importance, the latter being of special concern for defendants who are in jail. Apparently the prestige factor for defendants out of jail (attending trial of their own accord) dominates the race difference which would otherwise favor the whites.

[18]This is in contrast to "crimes of passion" such as murder, which are more friendly affairs. In Detroit in three-quarters of the murder cases, assailant and victim are relatives or friends.

[19]This becomes a special case of the phenomenon one might call cross-case bargaining, in which the outcome for case A is sacrificed somewhat by the prosecutor in order to obtain a more favorable outcome from case B, with D's contribution to the cause-- assuming he is case A --being his availability as a source of information which P deems helpful to bettering his outcome from case B. Although far less frequent than the bread-and-butter mode of bargaining described by this model (single case bargaining which is pareto optimal within the two-party community of P and D,) cross-case bargaining is perhaps the more familiar of the two due to the wide publicity surrounding the Watergate-related cases.

[20]This was because the blind draw system for assigning judges (and knowing one's place in the queue of the assigned judge) did not begin until after the period of this data.

[21]A number of the case files involved two or more defendants at the time of arrest. The status of the case was continually recorded (usually) for one or possibly more of the defendants who survived various sources of attrition, such as transferral to other courts or other counties or states, dropping of charges, guilty plea at arraignment, etcetera. Forthe first survivor who continued through the regular plea process, the records were generally quite thorough. However, it appeared that the records of the others--labeled here as "two-plus" defendants--sometimes were filled in with the same entries as those of the first defendant, particularly the entry for the probability of conviction, which was typically evaluated for the first defendant and then attributed to the others also. Since the evidence of guilt could and often did vary widely among defendants on a single case, as did other factors affecting outcome, it was surprising to find the probability of conviction to be identical for all defendants on the case. In view of the pattern for repetition, it was believed that inclusion of the "two-plus" defendants would add to the errors-in-variables problems already present (that due to the many prosecutors making the estimates and the limited set of discrete values possible for assignment) and so only first defendants were used for the estimation of the relative strength variable and regressions on the outcome function.

[22]The limited range for the probability estimates has two possible justifications. One is that the prosecutors who make the estimates are insupportably optimistic, and thus feel that even the worst case has better than an even chance of winning.

The alternative explanation lies with a perception of the prosecutorial stage of the criminal justice process as not only a

processor prior to trial, but as a subagency which receives cases sifted out in the preceding stages of the process. Thus P does not see cases in which there was no probable cause for arrest (although these would receive low probability estimates) nor does he see arrests not made because a police officer and potential defendant traded information instead.

Thus if one believes that the previous stages in the process are good sifting devices (or alternatively, that they are sufficiently desperate for funds that priorities are set), then it is a logical consequence that only the fittest of the potential cases survives to the point of entering the prosecutor's files. This is equivalent to a sort of bureaucratic Darwinism.

[23]As it turned out (because of the later institution of the blind draw system) this piece of information was not usually known by defensants until the pretrial conference, thus allowing scant time or no time at all for the implications of the judge assignment to be taken into account in the estimates of r. As revealed below, this variable wins the contest for being the least significant of the ones entered in the regression equation.

[24]Of course, q is not linear in the variables, since Race and Jail are entered in interaction with one another, as well as being entered separately.

[25]When run with all cases for which the necessary data existed (including two-plus defendants) the results were close to the same as those reported in the text:

$$\ln\left(\frac{1-r_p}{r_p}\right) = \underset{}{-1.37} \underset{(-3.24)}{-1.15\ \text{Jail}} \underset{(-2.11)}{-.62\ \text{Race}} + \underset{(2.17)}{.88\ \text{Race.Jail}}$$

$$+ \underset{(1.09)}{.002\ \text{Total time}} + \underset{(1.02)}{.01\ \text{J Plea}} - \underset{(-1.55)}{.298.\text{Pend Chg}}$$

$$+ \underset{(1.17)}{.32\ \text{Know Wit}} - \underset{(-1.02)}{.23\ \text{Age 17}} + e_i$$

$$R^2 = .1712 \qquad n = 109$$

Additional regressions are displayed in the data appendix.

[26]It would have been desirable to have information on whether the counsel for defendants had been privately retained or had been assigned by the court. However, several factors mitigate the need for such data. First, there are four sets of attorneys who represent criminal defendants: private attorneys who do not specialize in criminal law alone, attorneys who do specialize and are retained by defendants, private attorneys assigned to defendants and attorneys who practice through the

Defender's Association (which is funded in part by the Detroit Bar Association and the fees paid by the court.) Since judges assign cases to private attorneys who practice in the court anyway, the second and third groups are largely the same. The first group is small if it exists at all. The contrast which would be of interest is between the privately practicing attorneys and those from the Defender's office--and the distinction there is not one of the income of the defendant.

Second, the number of privately practicing attorneys is small; they are specialists and are physically present at the court much of the day (thus limiting time for preparation.) From the notes of the American Bar Foundation study, Michigan Field Reports (1958)

> "There are only seven or eight bona fide "criminal lawyers' in the city of Detroit...
>
> "The primary reason, according to [a lawyer], for the small number of criminal lawyers in Detroit is the fact that a general practitioner simply does not have the time to devote to criminal matters. Those attorneys not appearing in Recorder's Court every day are required to spend at least one half a day when the defendant is arraigned on the warrant. A full day is required at a preliminary examination, another half a day is required when a defendant is arraigned on the information. On the other hand, the criminal lawyers, being in Recorder's Court anyway, only consider a few minutes of their time being taken up on arraignments. Preliminary examinations can be conducted in a matter of 20 minutes or so. The civil lawyers, therefore, find it necessary to charge much higher fees than are required by the criminal lawyers. Since a vast majority of the defendants in Recorder's Court are not able to pay the fees required by the 'general practitioner' most of the criminal cases are handled by the few criminal lawyers."

Thus the absence of information on whether the defendant was assigned counsel is of relatively little harm as relates to the income of the defendant.

3.3 The Disposition

In the preceding two sections, the model was adapted and the data made compatible with the model for use in testing the outcome function determined by a proportional split of the Pareto optimal region between prosecutor and defendant. Substituting the assumed parameter values, and representing the h-functions as $h_p = -1/\hat{r}$ and $h_d = -\hat{r}$ the outcome function to be estimated is

$$v_m = -\lambda(.0065) \cdot \frac{R^{155}}{R'^{154}} \cdot \hat{r}^{-1} + (1-\lambda) \cdot \bar{v} + (1-\lambda) \cdot \gamma \cdot \underline{R} \cdot \hat{r} + \epsilon$$

which can be restated as

$$v_m = a_1 x_1 + a_2 x_2 + a_3 x_3 + \epsilon$$

where $a_1 = -\lambda$, $a_2 = 1-\lambda$, $a_3 = 1-\lambda$, and the x_i are the first, second, and third terms respectively, stripped of their a_i coefficients and ϵ is randomly distributed. We expect that a_2 will be a number between zero and unity, and that a_1 and a_3 will be numbers such that $\partial v_m / \partial \hat{r} > 0$. We would also expect the absolute values of a_1 and a_3 to be less than one, with their signs as noted, except that the scaling problem may distort the magnitudes, as discussed further below.

If one recalls (1) the set of assumed parameter values, i.e., regarding the prosecutor's switch function, that $w = 1$, $\alpha = 1/155$, $\beta = 154/155$ (reflecting the weighting of P's cases); and regarding the defendant's switch function, that $p_c = 1$; and (2) the assumptions made in the process of readying the data for use here, such as requiring $\underline{R}$ and $\underline{R}'$ to be equal to one if they would otherwise be less, adapting the Census poverty statistics for use as an income variable, and making other assumptions entailed in the determination of the costs of trial

and case maintenance; then it is clear that some of these assumptions bear the potential for affecting the results of any attempt to estimate the outcome function. The manner in which the h-functions are stated will also precipitate this danger. These regrettable but necessary assumptions--which affect the magnitudes of x_1, x_2, and x_3, and are therefore referred to as "the scaling problem"--wage warfare with the empirical results on two interrelated fronts: first, they can escalate or de-escalate the scale of the x_i terms (especially x_1 and x_3) and thus threaten the sign of $\partial v_m/\partial \hat{r}$, since the sign of the latter results from the addition of two terms, possibly of opposite sign; and second, they determine the value of $r''_{disc.}$ and thus through γ sever one regression regime from another.

The first liability--the general problem of variable scaling--holds little hope for rescue, other than further attempts to adjust the parameter values so that the resulting scales of the x_i seem "reasonable". The second one, that of compensating for scaling problems in determining the regression regime to which an individual should be assigned, holds out a bit more promise of a reprieve. The "cut variable" γ was defined by comparison of $\hat{r}''_o$ with $r''_{disc.}$, and only the second of these is affected by problems of scaling. [1] Instead of simply defining γ to be 1 when $\hat{r}''_o \leq r''_{disc.}$ and 0 otherwise, we define a new variable $\xi = \hat{r}''_o - r''_{disc.}$ and cut variables $\gamma_j = 1$ if $\xi \leq k_j$ and 0 otherwise. The outcome function $v_{mj} = a_{1j}x_1 + a_{2j}x_2 + a_{3j}x_{3j}$ can be estimated for several alternative values of k_j (where $x_{3j} = \gamma_j \underline{R}\hat{r}$), and the results of these several regressions may be compared with one another to determine the value of k_j which yields the best results, as indicated

by the values of the coefficient of determination and the significance of the regression coefficients. [2] Once the best value of k_j (among those chosen, as the cardinality of the set of k_j must necessarily be quite small) has been found, the difference between the best k_j and zero may give some indication of the scaling problem resulting from the assumptions regarding $\bar{v}$, $\underline{R}'$, and p_c, since the theory implies that when the correct values of those terms are used, the best cut point is the one for zero difference between $\hat{r}''_o$ and $r''_{disc.}$.

The result of the regressions for ten alternative cuts of the r" domain are given in Table 3, in which the dependent variable is the maximum sentence associated with the offense to which D pleaded guilty. To account for the possibility that the appropriate measure of outcome was the difference between the plea offense maximum and the maximum for the attempt charge corresponding to the original charge [3], the same regressions were also run with that difference as the dependent variable, and those results are given in Table 4.

For both versions of the regression equation (i.e. for both of the alternative dependent variables) the best results occur at $\xi = .6$, with the plea maximum yielding a slightly better value of R^2 than did the plea/attempt difference, and the preference reversed for coefficient significance. However, in both instances the coefficient of determination is quite close to one-half, and the t-values of all coefficients are greater than two.

The fact that some cuts of the domain--actually any cut--yield better results than the "no cut" case,would tend to confirm the theory that the defendant specializes in trial improvement or in consumption.

TABLE 3 Estimation of the outcome function (with no constant term)

Dependent variable is maximum sentence associated with guilty plea offense.

Cut	a_1	a_2	a_3	R^2
k_1=-.25	-.510 (-1.74)	.632 (9.33)	3.854 (.29)	.5002
k_2=0	-.510 (-1.71)	.632 (9.33)	3.854 (.29)	.5002
k_3=.25	-.546 (-1.77)	.637 (9.28)	2.748 (.53)	.5014
k_4=.50	-.792 (-1.98)	.670 (8.76)	3.396 (1.09)	.5067
k_5=.55	-1.191 (-2.59)	.723 (8.87)	5.923 (1.94)	.5213
k_6=.60	-1.314 (-2.66)	.738 (8.72)	6.282 (2.03)	.5233
k_7=.65	-1.268 (-2.22)	.727 (8.01)	5.288 (1.57)	.5141
k_8-.70	-1.16 (-1.58)	.707 (6.84)	3.932 (.99)	.5055
k_9=.80	-1.38 (-1.51)	.700 (7.27)	4.671 (1.02)	.5059
k_{10}: no cut	-.322 (-.21)	.634 (8.08)	-1.077 (-.11)	.4998

Note: The first and last of the k_j represent cuts extremely near and at the bounds of r, respectively.

Above, $x_1 = .0065R^{155}/R'^{154} \cdot r^{-1}$

$x_2 = \bar{v}$

$x_3 = \underline{R} \cdot \hat{r}$.

Table 4. Estimation of outcome function using the difference between the maximum sentence for the guilty plea offense and the maximum for attempt on the original charge as the dependent variable.

Cut	a_1	a_2	a_3	R^2	F
None	-.04199 (-.347)	.571 (7.055)	-2.812 (-.298)	.4323	31.60
k_1=-.25	-1.1316 (-3.687)	.562 (8.051)	6.234 (.449)	.4331	31.70
k_2=.00	-1.1316 (-3.687)	.562 (8.051)	6.234 (.449)	.4331	31.70
k_3=.25	-1.1849 (-3.730)	.570 (8.057)	4.144 (.774)	.4357	32.049
k_4=.50	-1.6078 (-3.946)	.627 (8.035)	5.724 (1.792)	.4529	34.348
k_5=.55	-2.1419 (-4.627)	.697 (8.492)	8.798 (2.858)	.4826	38.701
k_6=.60	-2.3203 (-4.667)	.719 (8.454)	9.306 (2.990)	.4861	39.39
k_7=.65	-2.2733 (-3.928)	.706 (7.660)	7.977 (2.336)	.4667	36.32
k_8=.70	-2.0573 (-2.737)	.670 (6.310)	5.629 (1.348)	.4444	33.19
k_9=.80	-1.7679 (-1.873)	.612 (6.130)	3.499 (.738)	.4354	32.00

Above, $x_1 = .0065R^{155}/R'^{154} \cdot \hat{r}^{-1}$

$x_2 = \bar{v}$

$x_3 = \underline{R} \cdot \hat{r}$

Worded in a more rigorous manner, had the "no cut" case yielded the best results, we would have had to conclude either that such specialization does not take place, or that at least for the sample individuals, the point of specialization reversal occurred outside the boundaries of the set of observed v. [4] In any event, although this result tends to be supportive of the theory, it cannot be viewed as the outcome of a particularly strong test of that theory.

Turning to the regressions for $\xi = .6$ we find that the sign and magnitude of a_2 are as predicted; if there are no scaling problems with $\bar{v}$ (or if one can be assured that they are minor) the conclusion is that the outcome of bargaining is comparatively closer to D's reservation price than it is to P's. In fact, the bargainable difference is split close to the ratio of seven to three in favor of the prosecutor. This poses an interesting contrast with the split of the total amount of v, labeled $\bar{v}$, when we broaden our focus beyond the Pareto optimal set. Recall that the average reduction in maximum sentence was to less than four-tenths of the original maximum. If the whole interval from zero to $\bar{v}$ were considered Pareto optimal, this major reduction in charge would seem to imply that the outcome generally falls closer to P's least desired outcome, leading us to believe that the defendant is the stronger party in bargaining. However, once the region of bargainable outcomes has been subjected to the condition of Pareto optimality, the assessment of bargaining skill reverses: within the region bounded by the reservation "prices" of the parties, the outcome falls closer to the D's reservation price, thus making P the stronger bargainer. [5]

Second, in considering $\partial v_m / \partial r$, which we predicted to be positive, it is necessary to engage in some arithmetic calculation in order to determine the sign which the regressions would yield. Recalling that

$$\partial v_m/\partial r = \lambda(.0065) \cdot \frac{R^{155}}{\bar{R}',154} r^{-2} + (1-\lambda)\gamma\underline{R}$$

we can write

$$\partial v_m/\partial r = -a_1 x_1 r^{-1} + a_3\gamma\underline{R} .$$

Since the partial is a function of r as well as of other variables, we would evaluate it at the means of the remaining variables; there is potentially a conceptual problem involved in basing an evaluation on the mean of γ, which being dichotomous, is not well represented by its mean. The problem is moot, however, since $-a_1$ and a_3 are positive (as predicted) for both versions of the regression, and so $\partial v_m/\partial r$ must be positive also, even when γ is assigned the alternative values of zero and unity. Note that so long as a_1 and a_3 bear the signs anticipated for them, scaling problems cannot affect the sign of $\partial v_m/\partial r$. It is when the signs of a_1 or a_3 fail to meet this precondition, as in the additional regressions in the appendix, that the scaling problem can actually become a problem. [6] Thus, with the h-functions used here, our predictions on the sign of $\partial v_m/\partial r$ are fulfilled also. [7]

In deriving the outcome function as a linear combination of the switch functions of the two parties (neither of which had a constant term) we found ourselves with a function having no additive constant term. To test whether that omission was in error, the same sets of regressions were run without suppression of the constant, with the results reported in Table 5. [8] As the reader can verify, the constant

Table 5. Estimation of outcome function with constant included, and guilty plea maximum sentence as dependent variable.

Cut	Constant	a_1	a_2	a_3	R^2	F
None	39.355 (.15)	-4.507 (-.16)	.636 (7.98)	-26.041 (-.16)	.4999	27.32
k_1=-.25	-.402 (-.03)	-.426 (-.13)	.633 (8.05)	3.754 (.27)	.5002	27.35
k_2=.00	-.402 (-.03)	-.426 (-.13)	.633 (8.05)	3.754 (.27)	.5002	27.35
k_3=.25	-2.512 (-.19)	-.024 (-.01)	.644 (7.98)	2.848 (.54)	.5015	27.50
k_4=.50	.490 (.03)	-.895 (-.28)	.669 (7.908)	3.409 (1.07)	.5067	28.07
k_5=.55	-2.949 (.20)	-.580 (.19)	.731 (7.96)	5.953 (1.93)	.5216	29.80
k_6=.60	-6.412 (-.43)	-.413 (-.00)	.758 (7.77)	6.495 (2.06)	.5244	30.13
k_7=.65	-1.412 (-.09)	-.973 (-.31)	.731 (7.36)	5.286 (1.56)	.5142	28.93
k_8=.70	-1.257 (-.08)	-.895 (-.28)	.711 (6.41)	3.925 (.98)	.5055	27.94
k_9=.80	-5.007 (-.33)	-.393 (-.12)	.718 (6.43)	4.995 (1.06)	.5065	28.05

where $x_1 = .0065R^{155}/R'^{154}\hat{r}_{.}^{-1}$

$x_2 = \bar{v}$

$x_3 = \underline{R}.\hat{r}$.

was not significant at any cut, and additionally reduced the reliability of the coefficients of the other terms, from which we can conclude that the version from the theory is the preferred form.

In summary, the form of the outcome function which provided the best results was the one which observed most strictly the form of the function constructed in the theory; the maximum sentence of the plea offense and the difference of that variable from the maximum for attempt on the original charge yielded similar results, both with values of R^2 close to .5 and significant t-values for the regression coefficients.

The question of whether the function estimated is in fact discontinuous at some point was tested by comparison of the discontinuous function at its best cut with the cut points at either end of the domain of $\hat{r}$; for the function descended from the theory, the best cut (in fact, any cut) was better than the alternative formulations, thus lending support to the belief that defendants specialize in trial outcome improvement or in consumption.

The sign of the partial derivative of outcome with respect to relative strength was positive, as predicted, and the coefficients of the term $\bar{v}$ indicated that the prosecutor generally drove a much harder bargain than did the defendant. It would appear that the estimation of the outcome function not only verifies the formulation of the model--subject to the assumptions made concerning parameter values and budget costs--but is also helpful in evaluating the results of negotiations in terms of the distribution of the bargainable amount of the good V between the two parties to the case.

Table 6. Proportion of cases in left-hand regime for different cuts of the domain.

No cut	.000
$k_1 = -.25$	.012
$k_2 = 0$	.012
$k_3 = .25$	.058
$k_4 = .50$	.291
$k_5 = .55$	.395
$k_6 = .60$	.442
$k_7 = .65$	.512
$k_8 = .70$	.616
$k_9 = .80$	.814

Notes to 3.3

[1]Recall that $r''_{disc.}$, the level of r" at which D would reverse his specialization was defined to be

$$r''_{disc} = \frac{[(\bar{v}+1)\ \ln(c'_o+1)+1]}{(\bar{v}+1)^2\ \ln(c'_o+1)} \cdot c_o$$

so that the scaling problem is introduced in two ways: (1) through the values of c_o and c'_o which are subject to assumptions about origins and amounts of D's resources; and (2) the estimates of time until trial.

[2]The modified search technique employed here is reminiscent of the work on time-varying parameters pioneered by Richard Quandt and developed further by David Belsley (1973), Alexander Sarris (1973), and others.

Quandt's approach was to cut the domain at each value of t, a discrete-time variable and to separately estimate a linear function on each regime thus defined. He then determined the best cut (value of t) to be the one which minimized

$$\lambda = \frac{\hat{\sigma}_1^t \hat{\sigma}_2^{T-t}}{\hat{\sigma}^T}$$

where $\hat{\sigma}_1$ and $\hat{\sigma}_2$ denote the standard errors of estimate of the left- and right-hand regimes, respectively, and $\hat{\sigma}$ for the whole domain of t. See Quandt (1958, 1960) and Belsley (1973) and Sarris (1973) papers.

[3]Walton and McKersie (1965) contend that once common points of settlement have been recognized ("once" referring to the historical pattern of negotiations, not an event within a single bargaining session), as the attempt charge may be regarded in Detroit, bargaining will often lead to that common point. If they are correct, one might want to consider the deviations from that common point, rather than simply the settlement point alone, as the relevant outcome measure.

[4]On the other hand, failure to reject the hypothesis that the discontinuity exists, as is the case here, can be taken to imply that the discontinuity does exist, or else that the introduction of the discontinuity allows a better fit of another function (even though the other function may be continuous.) For example, if the true function were the nonlinear function drawn below, and we used two linear ones to estimate, we should not conclude that the true function is discontinuous, just because the discontinuous pair describes the true function well:

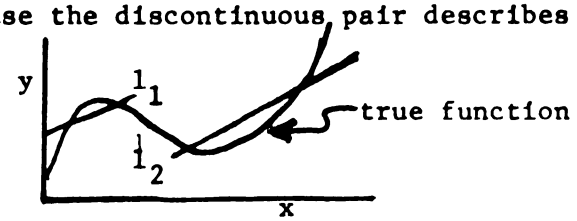


Figure 28. Example of a discontinuous function estimating well a continuous one.

[5]Comparison of the split of $\bar{v}$ with that of the pareto optimal set suggests another point as well. As discussed and sketched in sections 2.1 and 2.4, the pareto-optimal interval changes in both location and size, depending upon the level of $\underline{v}$; this implies that any assessment of bargain outcomes based upon $\bar{v}$ alone may not only be in error conceptually, but may be confirmed empirically if the observations tend to fall at one extreme of the domain of v. The construct tested here does not fall victim to that problem in part because the use of switch functions immediately restricts it to the pareto optimal set, and also because the proportion of the split is assumed to be fixed throughout the domain of v.

[6]In fact, even for the additional sums, which are displayed in the data appendix, the signs of $\partial v_m/\partial r$ are correct. The potential problem lies in the marriage of the scaling problem to that of the non-independence of the first and third variables: the latter may reverse one or more signs of coefficients; if that has occurred, then the scaling of the terms to be "added" (in fact, one subtracted from the other) becomes a crucial determinant of the sign of $\partial v_m/\partial r$.

[7]The fact that a_1 and a_3 do not lie in the intervals [-1,0] and [0,1], respectively, only serves to point out again the scaling problems inherent in the model as applied to this data.

[8]The regressions for the other possible cut points appear in the appendix.

4. Conclusions

In the beginning of this dissertation, various aspects of the judicial process were interpreted within an economic construct. Each of the two parties to negotiations, the prosecution and the defense, was believed to calculate the outcome and utility he anticipated from trial, and then the concession which would be required by his opponent in order to make the prospect of a plea of guilty equally attractive. For each party, the trial option was represented by one constraint and the guilty plea option by another. Thus the "switch function" which resulted from repeating the above calculation for alternative values of the independent variable was in effect the demand function for one good--a plea of guilty--stated in terms of the price of a substitute good, trial. Each opportunity set was a production-possibility frontier describing the case outcome in terms of variable inputs to trial (additional sleuthing and research effort) and the relative strength of the two parties. Concession was interpreted to be deliberate underutilization of the strength of one's own case.

The integration of the above items within the model led to two observable theoretical results. First, the derivation of the switch functions--each of which divided the strength-outcome space into a set of acceptable bargain outcomes and a set of unacceptable ones--lent a specific functional definition to the concept of a pareto optimal set of outcomes of negotiation. Hence, if one accepts the precept that

under any bargaining theory pareto optimality is prerequisite to a trial settlement, the set of negotiated outcomes should fall within the region bounded by the switch functions [1]. Due to the lack of appropriate statistical tools, this could only be "tested" by visual observation of data plots.

Second, a split-the-difference theory of bargained outcomes was posited, and the outcome function that suggested was a linear combination of the two switch functions derived earlier. Estimation of the outcome function served not only to test the general validity of the model, but also to determine the proportions in which the difference was split. In addition, since the defendant's switch function was expected in general to be discontinuous at some value of relative strength, due to his supposed specialization either in trial improvement or in consumption, the outcome function was expected to be discontinuous also (or at least nondifferentiable) at some point. Although not providing a particularly strong test of these contentions, since continuous functions are sometimes well estimated by discontinuous ones, a search for the point of discontinuity could serve to reject that hypothesis were the data to be best fitted by the continuous version of the outcome function, i.e. by a version in which the discontinuity fell outside the relevant portion of the domain.

The empirical results tended to support the premises of the theory--for example, that outcome level rose with the relative strength of the prosecutor--as well as to affirm the outcome function itself. The search for a point of discontinuity yielded positive results for all variants of the outcome function which were estimated, and the

split of the Pareto optimal set was found consistently to be one which favored the prosecutor.

In order to estimate outcome as a function of relative strength it was necessary to first create an index of relative strength based upon the factors which were believed to be critical to the determination of strength at trial. For that purpose estimated values of the prosecutor's estimate of the probability of conviction were used, where the estimates were the predicted values of that variable. Those values were obtained by regressing the probability estimates against a list of characteristics of the defendant and of the offense, and the initial endowment of evidence, where a logistic form was used for estimation. The signs of the coefficients were consistent with much of the interdisciplinary literature about the effects of those various factors upon the treatment of individuals within the criminal justice system.

Subject to the usual caveats concerning the assumptions made in developing the model and imperfections of the data, the results noted above have a number of implications both for further research and for policy formulation. The crucial factor of jail status permeates the model and is confirmed in the empirical work; even the strong effects of race in the determination of relative strength pale next to custody status. Further, the custody status of the defendant typically is crucial in terms of his ability to earn or receive income, suggesting that both policy and academic research consider not only the separate effects of such factors, but also their joint effects upon the outcome of alternative processes for resolution of the case. Thus, for example, current release-on-recognizance programs should be evaluated in terms

of their income (i.e. family welfare) and case-outcome effects as well as in terms of capiases issued (documents issued for failure to appear at trial.) Similarly, negative tax programs, unemployment compensation and other sources of income to defendants warrant consideration of other trial-related factors as well.

The assignment of income data based upon residential location poses an interesting problem in interpretation of the empirical results. In spite of the roughness of the data, and of the somewhat precarious manipulations made in deriving the level of disposable resources for individuals, the crosstabulations especially tended to link higher income figures with lower plea sentence maximums. This fact may be reflective of a tie between defendant resources and plea outcomes or it may be ghosting a relationship which is more complex. Since the Census entry used was the probability of being above the poverty line--and although we do not have data on individuals, criminal defendants are reputed to be rather homogeneous in income level--it is possible that this variable is in fact serving as a proxy for the social and economic amenities of the various neighborhoods in which defendants live and from which the Census data was drawn. In addition, this poses a number of questions about interrelations between income distribution and criminal activity. Clearly these questions offer a rich area for further research.

The time-honored maxim, justice delayed is justice denied, was borne out in the testing of the model, at least to the extent that the form of the outcome function (and thus the underlying opportunity sets) is affirmed. Since the time factor enters both through the resource

constraints and via the evidence effect of delay, analysis of different methods of rationing court time should take both into account. And again, because each of these effects is interwoven with others in the determination of the opportunity sets, there are a number of important relationships deserving of further investigation.

Although defendants obtain sizable reductions in charge for pleading guilty, when we consider the amount that is actually "up for grabs," that is, the Pareto optimal set, the prosecutor fares relatively better. The reasons for this--whether the administrative centralization peculiar to Detroit, the quality of defense attorneys, the resources of the parties relative to one another, or whatever--are deserving of more attention than their passing mention here.

The development and testing of a model such as this one bears implications for other areas of research (in particular, for application to bargaining theory) as well. The switch functions derived in the theory provide a two-dimensional generalization of the concept of the points of zero incremental utility for two bargaining parties. Second, the switch functions pose a concrete interpretation of the concept of a negotiated outcome which is exactly as attractive as its conflict substitute: the ability to derive such a point or function had always been assumed, but had not been demonstrated. The linking of strength in bargaining with a production process may shorten the distance between the deliberative, earthbound nature of traditional microeconomics and the ethereal aura of the bargaining theories, which assume that all issues of micro theory have been settled before making their appearance.

Although not a dynamic model of the bargaining process, the construct introduced here is consistent with the ideas of the negotiated outcome being the result of a series of concessions, one in response to another, and of there being a dominant pattern of bargained outcomes--as represented by the split of the Pareto optimal set--between the two parties to negotiation. Although the assumption that the outcome function is a linear combination of the switch functions is open to criticism, this same assumption assures that the two-party nature of the bargain process is not sacrificed, as has often been necessary in order to obtain a testable equation (deMenil 1971; Bental and Comay 1973; Comay, Melnick, and Subotnik 1974; Ashenfelter and Johnson 1969; and others.)

Finally, use of a split-the-difference theory suggests that Occum's Razor may be a useful tool in selecting a bargaining theory, and also adds to the recent store of empirical work in an area of economics long believed to hold little potential for such application.

Where to go from here? In addition to those cited above, some topics for additional research include incorporation in the model of uncertainty and the role of information; greater detail in the development of the production function for case outcomes; an economic theory about the determination of relative strength; the creation of models of other stages in the criminal justice system and their linkages with those of the courts; and the extension of the theory to an aggregate model of the market for guilty pleas. Coupled with data on these same issues, the existence of such theories would permit a panoply of addi-

tional questions to be broached concerning the process and incidence of plea bargaining and other legal processes.

APPENDICES

APPENDIX A.1

Data Summary Tables

Table 7. Recorder's Court Judges: Statistics Derived from 1972 Annual Reports

	Name	Percent Pleas	Percent Dismissals	J/W	J-W / J+W
1	Colombo	76	23	.86	-.08
2	Crockett	45	33	.23	-.63
3	Davenport	73	15	.73	-.16
4	Evans	64	29	.19	-.68
5	Ford	41	41	9.99*	.92
6	Gillis	53	47	3.00	.50
7	Heading	58	33	3.30	.53
8	Leonard	88	10	.78	-.13
9	Maher	79	11	1.81	.29
10	Murphy	72	27	.71	-.17
11	Olsen	78	16	.29	-.56
12	Poindexter	78	15	1.62	.24
13	Schemanske	58	16	6.97	.75
	Ave. Regular Judges	69	26	1.00	-.00(-.004)
	Ave. Visiting Judges	60	33	1.28	-.12
	Ave. Total Cases	64	30	1.15	+.17

* The true figure is 24.43
J denotes jury trial; W waiver of jury.

Table 8. Original Offenses Charged: Years Maximum Sentence

Years	Absolute Frequency	Relative Frequency (Percent)	Cumulative Adjusted Freq. (Percent)
2.00	1	0.6	0.6
2.50	4	2.6	3.2
4.00	7	4.5	7.7
5.00	56	36.1	43.9
10.00	32	20.6	64.5
15.00	20	12.9	77.4
40.00	35	22.6	100.0
Total	155	100.0	100.0

Mean	15.097
Median	8.984
Standard Error	1.119
Standard Deviation	13.937

Table 9. Guilty Plea Offenses: Years Maximum Sentence

Years	Absolute Frequency	Relative Frequency (Percent)	Cumulative Adjusted Freq. (Percent)
0.0	16	10.3	10.3
0.20	16	10.3	20.6
1.00	2	1.3	21.9
2.00	54	34.8	56.8
2.50	19	12.3	69.0
4.00	10	6.5	75.5
5.00	17	11.0	86.5
10.00	2	1.3	87.7
15.00	3	1.9	89.7
40.00	16	10.3	100.0
Total	155	100.0	100.0

Mean	6.392
Median	2.104
Standard Error	.939
Standard Deviation	11.696

Table 10. Jail Status of Defendants

	Value	Absolute Frequency	Relative Frequency (Percent)
Not in Jail	0.0	72	46.5
In Jail	1.00	83	53.5
	Total	155	100.0

Table 11. Race of Defendants

	Value	Absolute Frequency	Relative Frequency (Percent)
White	0.0	35	22.6
Nonwhite	1.00	120	77.4
	Total	155	100.0

Table 12. Race and Jail: Interaction Variable

Value	Absolute Frequency	Relative Frequency (Percent)
Nonwhite and in Jail	64	41.3
Others	91	58.7

Table 13. Time until Pretrial Conference and Total Time until Trial (in days)

	Until Pretrial Conference	Total Time Until Trial
Mean	70.11	142.15
Median	27.25	102.25
Standard Error	8.71	10.72
Standard Deviation	108.43	133.50

Table 14. Age of Defendants

Age	Absolute Frequency	Relative Frequency (Percent)	Cumulative Adjusted Freq. (Percent)
16	3	1.9	1.9
17	23	14.8	16.8
18	28	18.1	34.8
19	14	9.0	43.9
20	14	9.0	52.9
21	11	7.1	60.0
22	10	6.5	66.5
23	6	3.9	70.3
24	9	5.8	76.1
25	5	3.2	79.4
26	5	3.2	82.6
27	2	1.3	83.9
28	5	3.2	87.1
29	2	1.3	88.4
30	3	1.9	90.3
31-35	5	3.1	93.5
36-40	5	3.1	96.8
Over 40	5	3.1	100.0

Table 15. Defendants Having Other Cases Pending

Number of Other Cases	Absolute Frequency	Relative Frequency (Percent)	Cumulative Adjusted Freq. (Percent)
0.0	118	76.1	76.1
1.00	30	19.4	95.5
2.00	4	2.6	98.1
3.00	1	0.6	98.7
4.00	1	0.6	99.4
5.00	1	0.6	100.0
Total	155	100.0	100.0

Mean	.323
Median	0.0
Standard Deviation	.720

Table 16. Relationship of Defendant and Complainant

	Absolute Frequency	Relative Frequency (Percent)
Know Each Other (Family Friendship or Regular Contact)	18	11.6
Do Not Know Each Other	137	88.4

Table 17. Previous Arrests

	Absolute Frequency	Relative Frequency (Percent)
None	77	49.7
One or More	78	50.3

APPENDIX A.2

Selected Scatter Plots

The object of the first part of the theory developed here was to determine the functions which bound the set of Pareto optimal plea bargain outcomes. Since there do not exist statistical tools which test for achievement of that property in the switch functions derived, the following plots are offered for observation. Plea bargain outcomes, taken to be the maximum sentence on the pleaded offense, are graphed on the vertical axis against the estimated relative strength on the horizontal axis, where the latter increases with greater prosecution strength. This same plot appears for different subsets of the sample used in the regression analysis of section 3.3.

Figure 29. White defendants: relative strength and guilty plea maximum

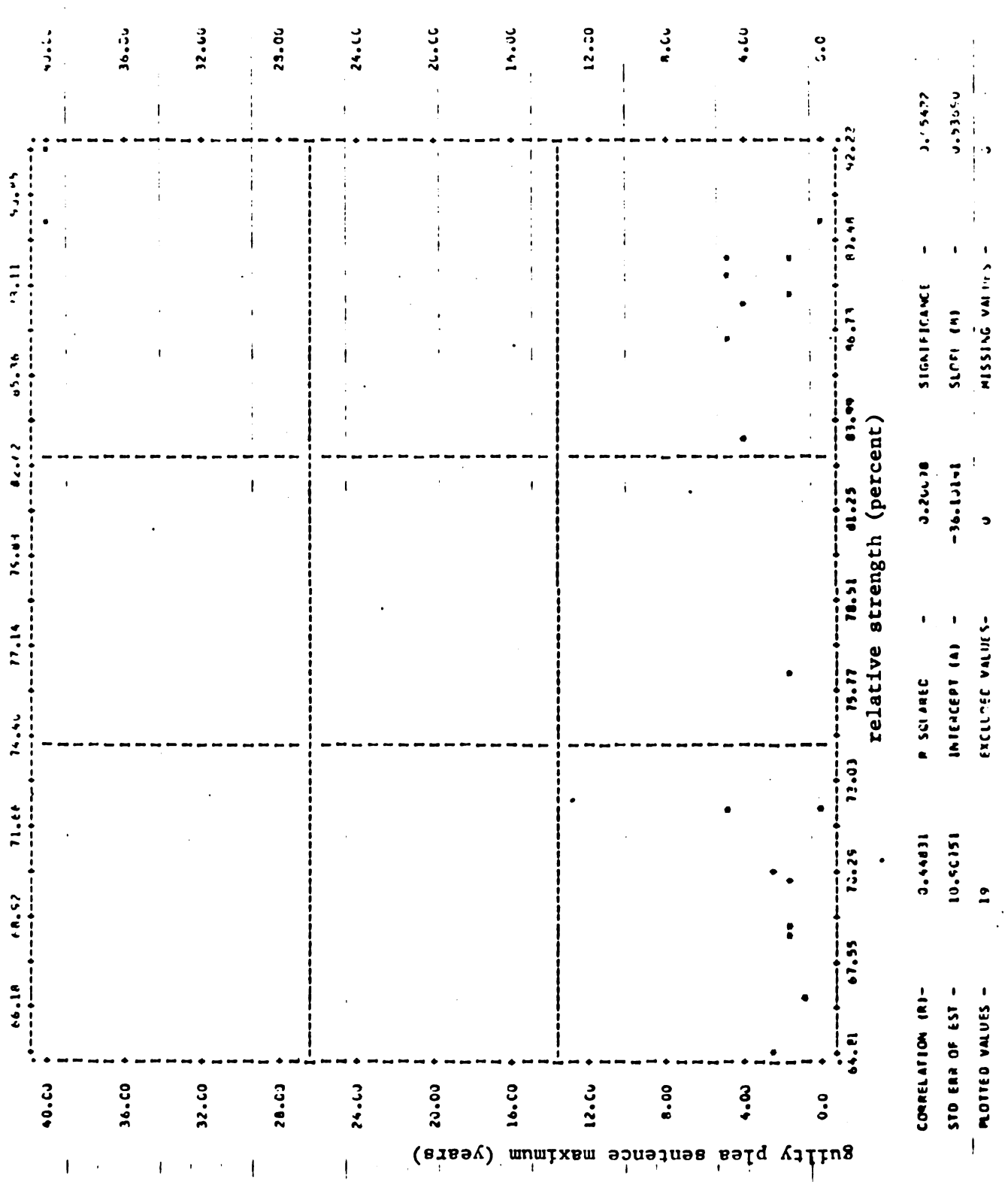


Figure 30. Nonwhite defendants: relative strength and guilty plea maximum

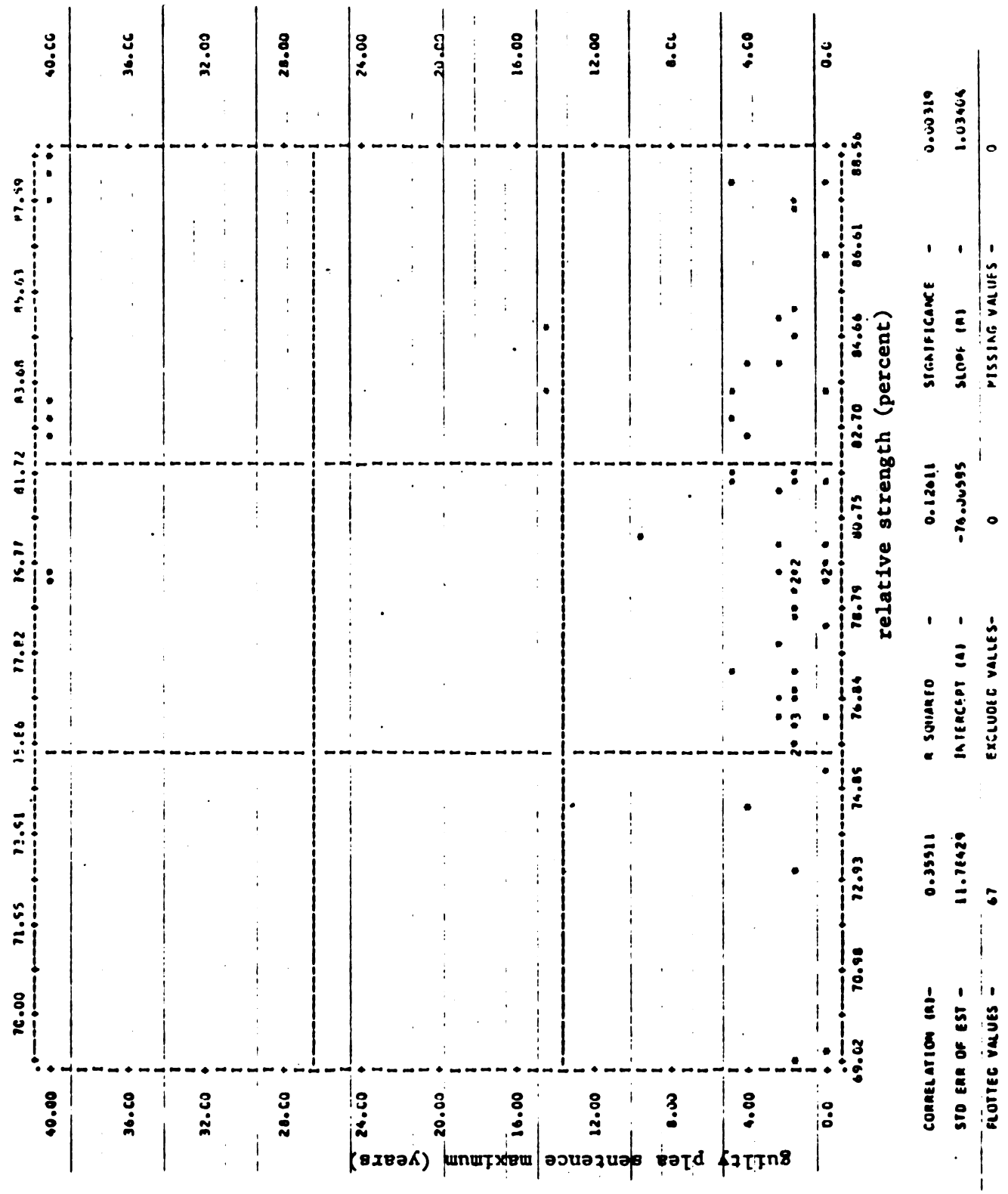


Figure 31. Defendants with arrest records: relative strength and guilty plea maximum

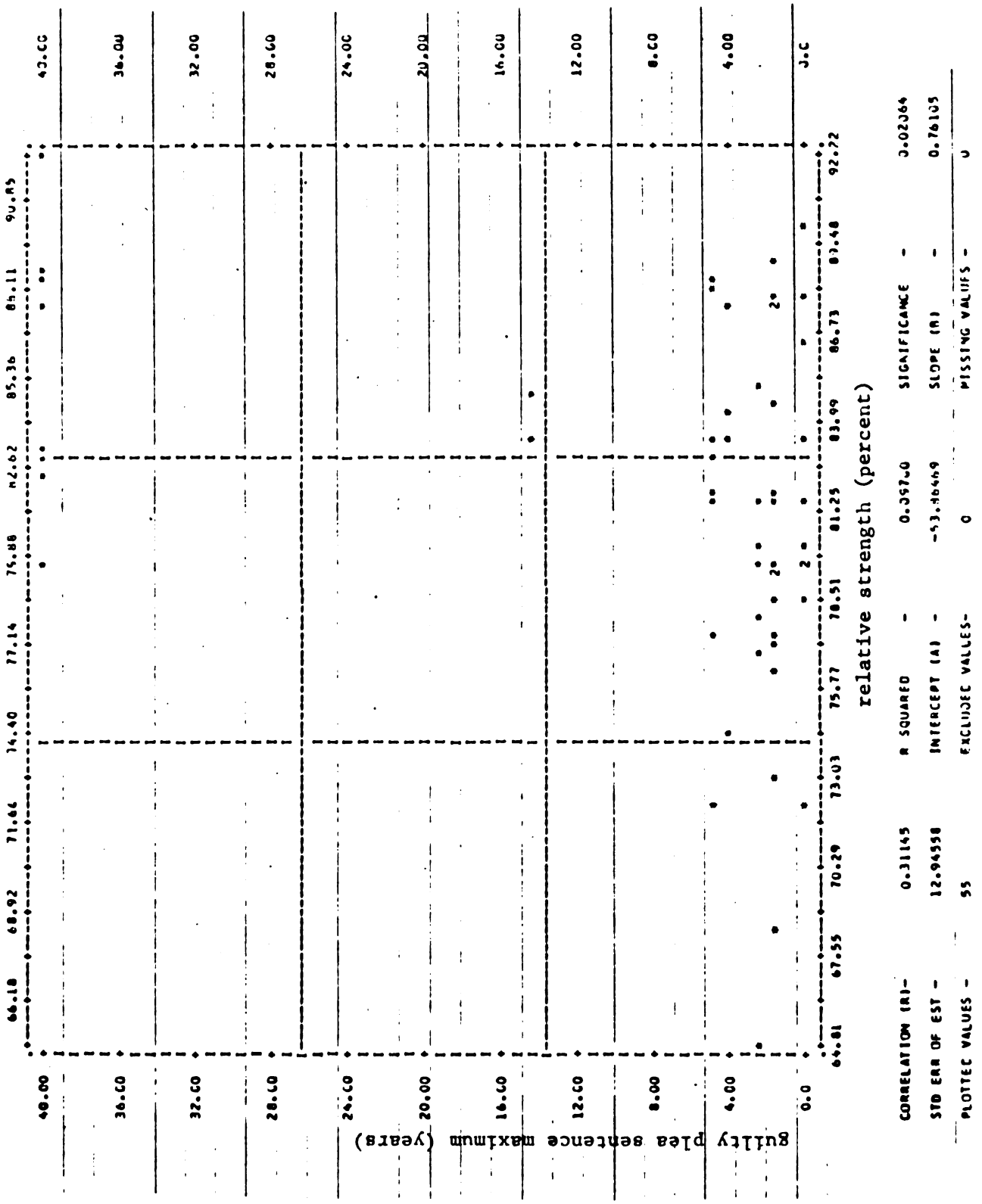
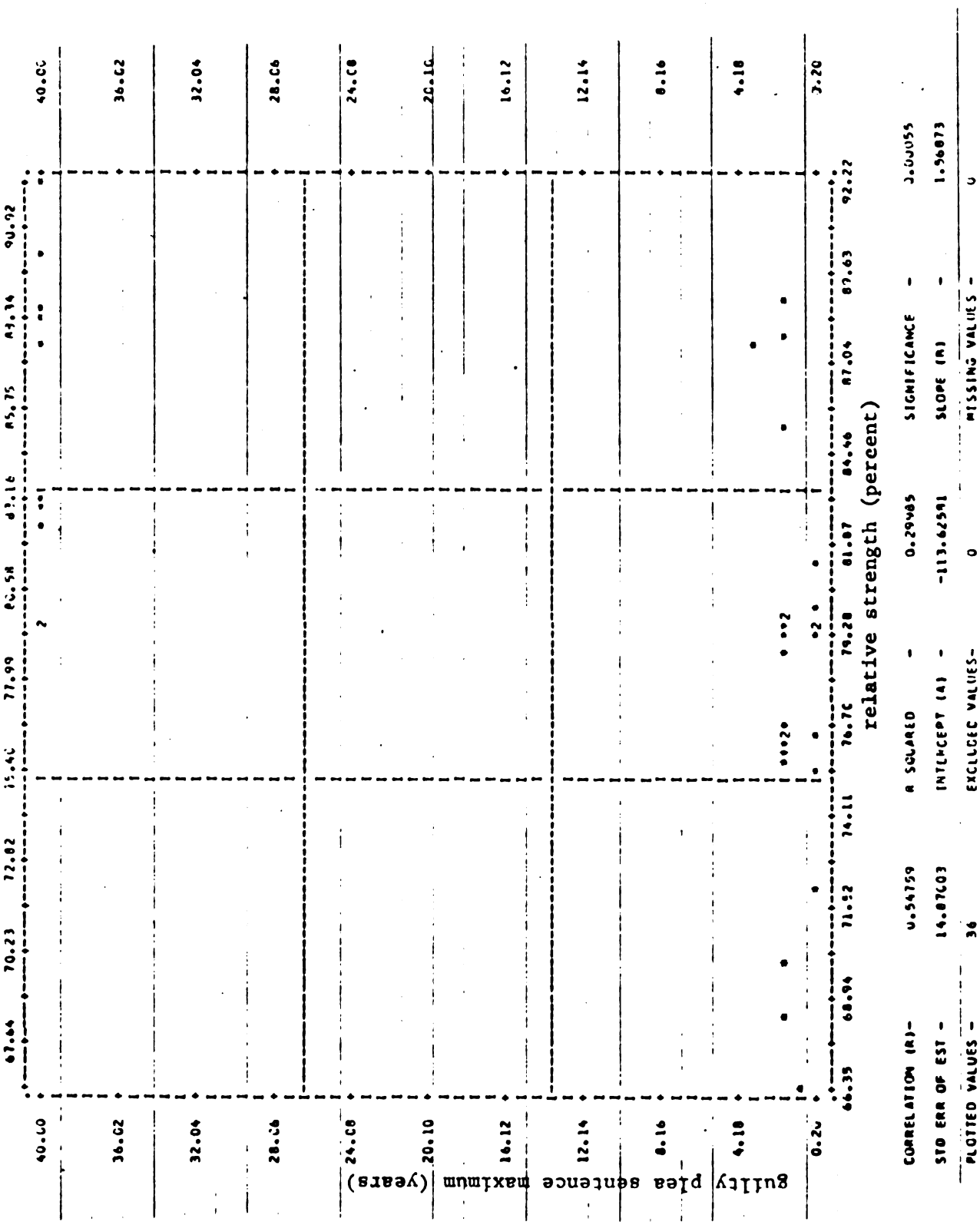


Figure 32. Defendants whose charges were reduced to a misdemeanor: relative strength and guilty plea maximum



APPENDIX A.3

Additional Regression Results

The first two sets of regression results display the alternative specifications of the linear function used in estimating relative strength from the data on the prosecutor's opinion about the probability of conviction at trial.

The latter two tables present the results of regressions run in the same manner as the ones shown in section 3.3, except that here the resources of both parties are assigned their mean values. Resources are held constant at those levels in order to reduce the amount of cross-variation within the first and third terms, which are each the product of several variables.

TABLE 18 Estimation of the Relationship of Relative Strength (Probability of Conviction) To Characteristics of the Defendant and of the Offense. Dependent Variable is the Natural Logarithm of the Odds in Favor of Defense

(First Defendants only are Included)

	Constant	Jail	Race	Race.Jail	Total Time	J.Plea	Pend.Chg.	Know.Wit	Age 17	Age 18	Age 24	R^2	Sig
Eq.1	-1.37	-1.22 (-3.47)	- .53 (-1.84)	1.03 (2.55)	.002 (1.11)	.005 (.69)	005 (-1.83)	.32 (1.21)		-.06 (-.34)		21.82	.012
Eq.2	-1.24	-1.24 (-3.55)	- .55) (-1.98)	1.05 (2.63)	.002 (1.07)	.004 (.56)	- .36 (-1.94)	.30 (1.16)	-.22 (-.98)			22.66	.009
Eq.3	-1.43	-1.22 (-3.45)	- .52 (-1.80)	1.02 (2.52)	.003 (1.12)	.006 (.71)	- .35 (-1.84)	.31 (1.45)				21.72	.012
Eq.4	-1.13	-1.24 (-3.49)	- .55 (-1.88)	1.05 (2.58)	.002 (1.04)	.004 (.54)	- .37 (-1.90)	.32 (1.15)	-.22 (-.96)			22.67	.016
Eq.5	-1.37	-1.22 (-3.44)	- .53 (-1.82)	1.03 (2.52)	.002 (1.09)	.005 (.68)	- .35 (-1.79)	.32 (1.17)		-.06 (.96)		21.82	.021
Eq.6	-1.38	-1.14 (-2.94)	- .50 (-1.71)	.96 (2.28)	.0025 (1.11)	.005 (.64)	- .32 (-1.55)	.31 (1.16)			-.004 (-.53)	22.01	.020
Eq.7	-1.11	-1.17 (-3.03)	- .53 (-1.83)	.997 (2.38)	.002 (1.05)	.004 (.50)	- .34 (-1.69)	.31 (1.17)	-.21 (.94)		-.003 (-.45)	22.86	.015
Eq.8	-1.34	-1.14 (-2.94)	- .51 (-1.72)	.97 (2.29)	.002 (1.09)	.005 (.62)	- .32 (-1.60)	.33 (1.22)		-.05 (-.28)	-.003 (-.47)	22.05	.019
Eq.9	-1.10	-1.17 (-3.01)	- .53 (-1.79)	.996 (2.36)	.002 (1.04)	.004 (.49)	- .34 (-1.63)	.31 (1.16)	-.21 (-.91)		-.003 (-.43)	22.86	.025
Eq.10	-1.35	-1.15 (-2.93)	- .51 (-1.71)	.97 (2.27)	.002 (1.09)	.005 (.63)	- .32 (-1.53)	.32 (1.17)		-.04 (-.21)	-.003 (-.48)	22.06	.033

TABLE 19 Estimation of the Relationship of Relative Strength (Probability of Conviction) To Characteristics of the Defendant and of the Offense. Dependent Variable is the Natural Logarithm of the Odds in Favor of Defense

("Two-plus" Defendants are Included)

Constant	Jail	Race	Race.Jail	All Time	J.Plea	Pend.Chg.	Know.Wit	Weapon	Age	Age 17	Age 18	R^2	Sig
-1.79	-1.06 (-2.57)	- .54 (-1.71)	.78 (1.79)	.002 (1.01)	.01 (1.17)	- .28 (-1.42)	.33 (1.20)	-.03 (-.13)	.004 (.31)			16.34	.032
-1.37	-1.16 (2.94)	- .62 (-2.10)	.89 (2.12)	.002 (1.08)	.01 (1.02)	- .30 (-1.51)	.32 (1.16)	.01 (.04)		- .23 (-1.02)		17.13	.023
-1.51	-1.14 (-2.90)	- .63 (-2.11)	.88 (2.10)	.002 (1.14)	.01 (1.09)	- .27 (-1.34)	.35 (1.26)	.001 (.0053)			-.16 (-.89)	16.93	.025
-1.79	-1.08 (-2.91)	- .55 (-1.74)	.79 (1.88)	.002 (1.04)	.01 (1.20)	- .289 (-1.55)	.33 (1.20)		.004 (.30)			16.33	.019
-1.37	-1.15 (-3.24)	- .62 (-2.11)	.88 (2.17)	.002 (1.09)	.01 (1.02)	- .298 (-1.55)	.32 (1.17)			- .23 (-1.02)		17.12	.013
-1.51	-1.14 (-3.21)	- .63 (-2.13)	.88 (2.16)	.002 (1.15)	.01 (1.11)	- .267 (-1.39)	.35 (1.26)				-.16 (-.90)	16.92	.014
-1.77	-1.10 (-2.65)	- .54 (-1.70)	.78 (1.78)	.002 (1.08)	.01 (1.19)	- .26 (-1.30)		-.007 (-.03)	.0035 (.25)			15.13	.031
-1.37	-1.19 (-3.01)	- .61 (-2.07)	.88 (2.10)	.002 (1.15)	.01 (1.05)	- .275 (-1.39)		.03 (.13)		-.23 (-1.04)		15.99	.022
-1.54	-1.17 (-2.96)	- .62 (-2.06)	.86 (2.06)	.002 (1.21)	.01 (1.13)	- .24 (-1.22)		.02 (.09)			-.14 (-.79)	15.60	.026
-1.78	-1.10 (-2.96)	- .54 (-1.72)	.78 (1.84)	.002 (1.11)	.009 (1.21)	- .26 (-1.35)			.003 (.25)			15.30	.018
-1.36	-1.17 (-3.28)	- .61 (-2.08)	.87 (2.13)	.002 (1.15)	.009 (1.04)	- .27 (1.41)				-.23 (-1.04)		15.98	.012
-1.53	-1.15 (-3.24)	- .61 (-2.07)	.86 (2.10)	.002 (1.21)	.009 (1.14)	- .24 (-1.24)					-.14 (.79)	15.60	.014

TABLE 20 Estimation of Guilty Plea Maximum Sentence with Resources of Both Parites Held Constant

Cut	a_1	a_2	a_3	R^2	F
None	-.290 (-.19)	.635 (8.11)	-1.270 (-.13)	.4997	41.45
k_1=.25	-.509 (-1.71)	.632 (9.32)	3.867 (.29)	.5001	41.52
k_2=0.0	-.509 (-1.71)	.632 (9.32)	3.867 (.29)	.5001	41.52
k_3=.25	-.546 (-1.77)	.637 (9.27)	2.758 (.53)	.5013	41.71
k_4=.50	-.790 (-1.98)	.670 (8.75)	3.389 (1.08)	.5066	42.60
k_5=.55	-1.190 (-2.59)	.723 (8.86)	5.918 (1.93)	.5212	45.17
k_6=.60	-1.312 (-2.65)	.738 (8.71)	6.273 (2.02)	.5231	45.53
k_7=.65	-1.265 (-2.22)	.727 (7.99)	5.275 (1.57)	.5140	43.88
k_8=.70	-1.151 (-1.57)	.707 (6.83)	3.895 (.98)	.5053	42.39
k_9=.80	-1.369 (-1.50)	.700 (7.25)	4.637 (1.01)	.5057	42.46

where $x_1 = .0065R^{155}/R'^{154}.r^{-1}$

$x_2 = \bar{v}$

$x_3 = \underline{R}.Y$

and where R, R', and $\underline{R}$ are held constant at their respective means.

TABLE 21 Estimates of Difference Between Guilty Plea and Attempt Charge Maximum Sentences, with Resources of Both Parties held Constant

Cut	a_1	a_2	a_3	R_2	F (all <.0005)
None	-0.602 (-.37)	.572 (7.08)	-3.133 (-.32)	.4321	31.58
k_1=-.25	-1.131 (-3.68)	.562 (8.05)	6.272 (.45)	.4328	31.67
k_2=0.0	-1.131 (-3.68)	.562 (8.05)	6.272 (.45)	.4328	31.67
k_3=.25	-1.185 (3.73)	.570 (8.05)	4.171 (.78)	.4356	32.03
k_4=.50	-1.606 (-3.94)	.627 (8.03)	5.719 (1.79)	.4526	34.31
k_5=.55	-2.141 (-4.62)	.697 (8.48)	8.795 (2.85)	.4823	38.66
k_6=.60	-2.319 (-4.66)	.719 (8.44)	9.299 (2.98)	.4865	39.33
k_7=.65	-2.271 (-3.92)	.706 (7.65)	7.965 (2.33)	.4663	36.26
k_8=.70	-2.048 (-2.72)	.669 (6.30)	5.582 (1.37)	.4439	33.13
k_9=.80	-1.758 (-1.86)	.611 (6.11)	3.453 (.73)	.4350	31.96

where $x_1 = .0065R^{155}/R'^{154}.r^{-1}$

$x_2 = \bar{v}$

$x_3 = \underline{R}.Y$

and where R, R', and $\underline{R}$ are held constant at their respective means.

REFERENCES

REFERENCES

Abraham, Henry J. The Judicial Process, Second edition. New York: Oxford University Press, 1968.

American Bar Association, Advisory Committee on the Criminal Trial Standards Relating to Pleas of Guilty. New York: American Bar Association, 1967.

American Bar Foundation. Michigan Field Reports (Daily notes of investigators for the 1958 Study of Criminal Justice Systems.) Chicago: Library of the American Bar Foundation.

Ashenfelter, Orley A. and Johnson, George E. "Bargaining Theory, Trade Unions, and Industrial Strike Activity." A.E.R. 59 (March 1969): 35-49.

Bartholemew, Peter Prosecutor, Pre-Trial Division, Recorder's Court, Detroit. Interviewed July 31 and August 1, 1973.

Becker, Gary "Crime and Punishment: An Economic Approach." J.P.E. 76 (March - April 1968): 169-217.

_____. "A Theory of the Allocation of Time," Economic Journal 75 (September 1965): 493-517.

Belsley, David "On the Determination of Systematic Parameter Variation in the Linear Regression Model." National Bureau of Economic Research Report No. W0005. Cambridge, Mass., 1973.

Bental, Benjamin and Comay, Yochanan "A Dynamic Bargaining Model." Presented at the Winter Meetings of the Econometric Society (by Bental). Mimeo, October 1973.

Bishop, Robert L. "Game Theoretic Analyses of Bargaining." Q.J.E. 77 (November 1963): 559-602.

_____. "A Zeuthen-Hicks Theory of Bargaining." Econometrica 32 (July 1964): 410-417.

Boulding, Kenneth Conflict and Defense: A General Theory. New York: Harper, 1962.

Brady v. United States 397 U.S. 742 (1970)

Cartter, Allan M. Theory of Wages and Employment. Homewood, Illinois: Irwin, 1959.

Chamberlain, Neil Collective Bargaining. New York: McGraw-Hill Book Co., 1951.

Coase, R.H. "The Problem of Social Cost." J. Law Econ. 3 (October 1960): 1-44.

Coddington, Alan "A Theory of the Bargaining Process: Comment." A.E.R. 56 (June 1966): 522-530.

_____. Review of The Economoics of Bargaining by John G. Cross, J. Econ. Lit. 8 (1970): 1211-1212.

Comay, Yochanan; Melnick, Arie; and Subotnik, Abraham "Bargaining, Yield Curves, and Wage Settlements: An Empirical Analysis." J.P.E. 82 (March - April 1974): 303-313.

Cross, John G. The Economics of Bargaining. New York: Basic Books, Inc., 1969.

_____. "A Theory of the Bargaining Process." A.E.R. 55 (March 1965): 67-94.

_____. "A Theory of the Bargaining Process: Reply." A.E.R. 56 (June 1966): 530-533.

Cummings, L.L. and Harnett, D.L. "Bargaining Behavior in a Symmetric Bargaining Trial." R.E. Stud. 36 (October 1969): 485-502.

Dahl, Robert "The Concept of Power." Behavioral Science 2 (1957): 201-222.

Danziger, Sheldon and Wheeler, David "Malevolent Interdependence, Inequality, and Crime." Mimeo, M.I.T., September 1973.

Dawson, Robert O. Sentencing: The Decision as to Type, Length, and Conditions of Sentence. Boston: Little, Brown and Co., 1969.

de Menil, George Bargaining: Monopoly Power versus Union Power. Cambridge, Mass.: M.I.T. Press, 1971.

Douglas, Ann Industrial Peacemaking. New York: Columbia University Press, 1962.

Draper, N.R. and Smith, H. Applied Regression Analysis. New York: Wiley, 1966.

Dunlop, John Wage Determination Under Trade Unions. New York: Macmillan, 1944.

Farrell, M.J. "The Measurement of Productive Efficiency." Journal of the Royal Statistical Society 120 (Part III, 1957): 253-290.

Ferguson, Charles E. Microeconomic Theory, Third edition. Homewood, Ill.: Richard D. Irwin, Inc., 1972.

Foldes, Lucien "A Determinate Model of Bilateral Monopoly." Economica 31 (February 1964): 117-131.

Friedman, J.W. "On Experimental Research in Oligopoly." R.E. Stud. 36 (October 1969): 399-416.

Green, Edward Judicial Attitudes in Sentencing, A Study of the Factors Underlying the Sentencing Practice of the Criminal Court of Philadelphia. London: Macmillan & Co., Ltd., 1961.

Hadar, Josef Mathematical Theory of Economic Behavior Reading, Mass: Addison-Wesley Publishing Co., 1971.

Hamermesh, Daniel S. "Who 'Wins' in Wage Bargaining?" Industrial and Labor Relations Review 26 (July 1973): 1146-1149.

Harnett, Donald L. and Hamner, W. Clay "The Value of Information in Bargaining." Western Econ. J. 11 (March 1973): 81-88.

Harsanyi, John C. "Approaches to the Bargaining Problem Before and After the Theory of Games: A Critical Discussion of Zeuthen's, Hicks', and Nash's Theories." Econometrica 24 (April 1956): 144-157.

Harsanyi, John C. "Measurement of Social Power, Opportunity Costs, and the Theory of Two-Person Bargaining Games." Behavioral Science 7 (1962): 67-80.

Harsanyi, John C. "Notes on the Bargaining Problem." Southern Econ. J. 24 (April 1958): 471-482.

Hicks, John R. The Theory of Wages, Second Edition. London: Macmillan Co., 1963.

Johnston, J. Econometric Methods, Second edition. New York: McGraw-Hill, 1972.

Joseph, M.L. and Willis, R.H. "An Experimental Analog to Two-party Bargaining." Behavioral Science 8 (No. 2, 1963): 117-127.

Kasimar, M. and Choper, "The Right to Counsel in Minnesota: Some Field Findings and Legal-Policy Observations." Minn. L. Rev. 48 (1963): 41-44.

Katzman, Martin "The Economics of Defense Against Crime in the Streets." Land Economics 44 (November 1968): 431-440.

Klonoski, James R. and Mendelsohn,Allan "The Allocation of Justice: A Political Approach." pp.1-19 in Klonoski and Mendelsohn, eds. The Politics of Local Justice. Boston: Little, Brown and Co. 1970.

Kmenta, Jan Elements of Econometrics. New York: Macmillan, 1971.

Kraus, Alan and Melnik, Arie "A Sequential Decision Model of Bargaining Theory." Western Econ. J. 10 (December 1972): 359-375.

LaFave, Wayne Arrest: The Decision to Take a Suspect into Custody. Boston: Little, Brown and Co., 1965.

Landes, William M. "An Economic Analysis of the Courts." J. Law Econ. 14 (April 1971): 61-107.

_____, "The Bail System: An Economic Approach." J. Legal Stud. 2(January, 1973): 79-106.

Laster, Clarence, Jr. "Pretrial Conference Division (Explanation of Precedure's [sic] and Forms)" Mimeo, Pretrial Conference Division, Wayne County, Michigan, undated (post 1970).

Luce, R. Duncan and Raiffa, Howard Games and Decisions. New York: Wiley and Sons, 1957.

McConnell,John Chief, Pre-Trial Division, Recorder's Court, Detroit. Interviewed March 24, 1973.

McCoy v. United States 363 F. 2d, 306 (D.C. Cir. 1966).

McIntyre, Donald M., ed. Law Enforcement in the Metropolis, A Working Paper on the Criminal Law System in Detroit. Chicago: The American Bar Foundation, 1967.

Mendelsohn, Allan I. "The Influence of the Defendant's Plea on Judicial Determination of Sentence." Yale L. J. 66 (No.2, 1966): 204-222.

Miller, Frank W. Prosecution: The Decision to Charge a Suspect With a Crime, The Report of the American Bar Foundation's Survey of the Administration of Criminal Justice in the United States. Boston: Little, Brown and Co., 1969.

Mills, James "I Have Nothing to Do With Justice." Life (March 12, 1971): 56-67.

Morse, Wayne and Beattie, Ronald "Survey of the Administration of Criminal Justice in Oregon." Oregon L. Rev. 11 Supp. (1932).

Nash, John F., Jr. "The Bargaining Problem." Econometrica 18 (April 1950): 155-162.

_____, "Two Person Cooperative Games." Econometrica 21 (January 1953): 128-140.

Newman, Donald J. "Pleading Guilty for Consideration: A Study of Bargain Justice." J. Crim. L. C. and P.S. 46 (March - April 1956): 780-790.

_____. Conviction: The Determination of Guilt or Innocence Without Trial. Boston: Little, Brown and Co., 1966.

"Order in Court." Time 103 (April 1, 1974): 64-66.

Patchen, Martin "Models of Cooperation and Conflict: A Critical Review." J. Conflict Res. 14 (September 1970): 389-407.

Pen, Jan "A General Theory of Bargaining." A.E.R. 42 (March 1952): 24-42.

_____, The Wage Rate Under Collective Bargaining. (trans. by T.S. Preston) Cambridge, Mass.: M.I.T. Press, 1966.

Pencavel, J.H. "An Investigation into Industrial Strike Activity in Britain." Economica 37 (August 1970): 239-256.

People v. West, 3 Cal, 3d 595, 91 Cal. Rptr. 385, 477 P. 2d 409 (1970)

Peters, Edward Strategy and Tactics in Labor Negotiations. New London: National Foremen's Institute, 1955.

Posner, Richard A. Economic Analysis of Law. Boston: Little, Brown and Co., 1972.

Quandt, Richard E. "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes." J.A.S.A. 53 (1958): 873-880.

_____, "Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes." J.A.S.A. 55 (1960): 324-330.

The Recorder's Court of the City of Detroit, Annual Report. Detroit: County of Wayne, 1970, 1971, 1972.

Remington, Frank and others Criminal Justice Administration: Materials and Cases. Indianapolis: Bobbs-Merrill Co. Inc., 1969.

Remington, Frank and others 1972 Supplement, Criminal Justice Administration. Indianapolis: Bobbs-Merrill Co. Inc., 1972

Santobello v. New York 404 U.S. 257 (1971).

Saraydar, Edward "A Certainty-Equivalent Model of Bargaining." J. Conflict Resolution 15 (September 1971): 281-297.

_____, "Uncertainty, Bargaining, and the Strategy of Commitment." Zeitschrift fur Nationalokonomie 31 (1971): 309-322. (Also University of Western Ontario Department of Economics Reprint No. 95.)

Sarris, Alexander "A Bayesian Approach to Estimation of Time-Varying Regression Coefficients." National Bureau of Economic Research Report No. W0013. Cambridge, Mass., 1973.

Schackle, George L. S. "The Nature of the Bargaining Process." In John T. Dunlop (ed.) The Theory of Wage Determination; Proceedings of a Conference Held by the International Economic Association. London: Macmillan Co., 292-314.

Schaenman, Phillip Interviewed February 8, 1973 at the Urban Institute, Washington, D.C.

Schelling, Thomas C. The Strategy of Conflict. Cambridge, Mass.: Harvard University Press, 1960.

Shin, H.J. "Analysis of Charge Reduction And Its Outcome." Doctoral dissertation, State University of New York, Albany, 1972.

Shubik, Martin, Ed. Game Theory and Related Approaches to Social Behavior. New York: Wiley & Sons, 1964.

Siegal, Sidney and Fouraker, Lawrence E. Bargaining and Group Decision Making. New York: McGraw-Hill Co., 1960.

Silverstein, Lee Defense of the Poor in Criminal Cases in American State Courts. Berkeley, California: , 1965.

Thurow, Lester C. "Equity Versus Efficiency in Law Enforcement." Public Policy 18 (Summer 1970): 451-462.

Tobin, James "Liquidity Preference as Behavior Towards Risk." R.E. Stud. 67 (February 1958): 65-86. Also in Risk Aversion and Portfolio Choice ed. by Donald J. Hester and James Tobin, New Haven: Yale University Press, 1967.

Trebach, Arnold S. The Rationing of Justice. New Brunswick, N.J.: Rutgers University Press, 1964.

U.S. Census, Census of the Population, Fourth Count Tapes, 1970.

U.S. Constitution.

U.S. President's Commission on Law Enforcement and Administration of Justice The Challenge of Crime in a Free Society, Washington: U.S. Government Printing Office, 1967.

_____, Task Force Reports: The Courts. Washington: U.S. Government Printing Office, 1968.

United States v. Wiley 184 F. Supp. 679 (N.D. Ill. 1960).

Vetri, Dominick R. "Guilty Plea Bargaining: Compromises by Prosecutors to Secure Guilty Pleas." U. of Penn. L. Rev. 112 (1964): 865-908.

Virtue, Maxine Survey of Metropolitan Courts, Final Report. Ann Arbor, Michigan: University of Michigan Press, 1962.

von Neumann, John and Morganstern, Oskar Theory of Games and Economic Behavior, Third Edition. Princeton, N.J.: Princeton University Press, 1953.

Walton, Richard E. and McKersie, Robert B. A Behavioral Theory of Labor Negotiations: An Analysis of a Social Interaction System. New York: McGraw-Hill Co., 1965.

Wayne County Prosecutor's Office. Pretrial docket sheets and case files drawn from February to August, 1972. Detroit: Frank Murphy Hall of Justice (by permission).

Zeisel, Hans "Methodological Problems in Studies of Sentencing." Law and Society Review 3 (May 1969): 621-631.

Zeuthen, Frederik Problems of Monopoly and Economic Warfare. London: George Routledge and Sons, 1930: 104-150 (Chapter 4).