Game Theory and Literature*

STEVEN J. BRAMS

Department of Politics, New York University, New York, New York 10003

Received February 4, 1992

Applications of game theory to novels, short stories, plays, opera librettos, narrative poems, and the Hebrew Bible are surveyed from both an historical and a critical perspective, based in part on responses to a questionnaire. While some analyses shed light on literary issues, such as the role of emotions or the rationality of character choices that culminate in tragedy, others highlight game-theoretic issues, such as problems of coordinating choices or building reputations when information is incomplete. Several models indicate a sophisticated understanding of plots and character motives, but others are quite trivial or misuse game theory. Fiction writers, too, vary in the intuitive understanding of game theory that they bring to their works. Journal of Economic Literature Classification Number: 026. © 1994 Academic Press, Inc.

I. INTRODUCTION

Applications of game theory to the humanities have grown over the years in such disciplines as history, philosophy, religion, and linguistics. Ethical issues in business and law (e.g., related to fair representation) have also been analyzed using game-theoretic models. It is fiction, however, that has proved the most fertile ground for humanistic applications of game theory. Novels, short stories, plays, opera librettos, a narrative poem—all have been subject to game-theoretic exegesis, as have stories in the Hebrew Bible (Old Testament).1 It is these applications, most of which involve noncooperative game theory, that I shall survey here.

* I am indebted to Eva Brams, Peter Landweber, and an anonymous referee for valuable comments on an earlier draft. I thank all the game theorists who so generously responded to my questionnaire. The support of the C. V. Starr Center for Applied Economics is gratefully acknowledged.

1 But I know of no applications of game theory to art or music, including any to the performing arts.

0899-8256/94 $5.00
Copyright © 1994 by Academic Press, Inc.
All rights of reproduction in any form reserved.
I shall sketch but not present technical details of the models used, in part for lack of space and in part because my primary purpose is to emphasize literary themes amenable to game-theoretic treatment, not tools of the game-theoretic trade (on which game theorists need no instruction). This survey is meant to be reasonably comprehensive, but it is only a survey: those interested in the modeling details will have to go to the sources cited, which I hope to encourage by provoking interest in literature as a fruitful source of ideas for the game theorist.

Game theory may also provide a parsimonious framework and an important set of tools for the literary analyst. Although there are no rigorous tests to determine what the "right" interpretation of a work of fiction is, some interpretations are clearly more tenable than others. Game theory, in my opinion, has proved useful in explicating the strategic choices of characters by making tighter the linkage between motives and actions in plot construction. It is also useful in addressing certain interpretive questions, such as whether the ordinary calculations of fictional characters can help to explain their extraordinary actions in some of literature's great tragedies.

My review of applications of game theory to literature has both a critical and a historical dimension. In an attempt to gain an understanding of how and why the applications evolved as they did, I asked people who have applied game theory to fiction several questions, given in Section 2, where I also provide a chronological listing of the applications that have been made. In later sections, I make use of the respondents' answers to see both what inspired them to tackle a particular literary work and what they think a game-theoretic perspective brings to the understanding and interpretation of that work—and whether this work, in turn, stimulates the game theorist to probe new theoretical questions.

In discussing fictional works analyzed by game theorists, I begin in Section 4 by showing how two authors (Conan Doyle and Poe), instead of confronting the consequences of the Minimax Theorem in their fiction, sidestepped them. I then present in some detail a new application that illustrates how one writer (Faulkner) captured the spirit of the theorem, even invoking a fictitious "Player" to make seemingly random choices.

In Section 4, problems of coalition formation in constant-sum games take center stage in one play (by Pinter) and pervade one "political" novel (by Snow), arguably to the detriment of character development. One analyst, in fact, contends that emotions tend to be submerged when there are clear-cut winners and losers, whereas ambivalence is better expressed in literary plots with nonconstant-sum elements.

Several works of fiction that may be interpreted as nonconstant-sum games are reviewed in Sections 5 and 6—some quite critically, because of what I believe are some misuses of game theory in these applications.
Contrary to the views of some, I argue in Section 5 that great tragedies—like Shakespeare’s Othello, Puccini’s Tosca, and Shakespeare’s Richard III (usually classified as a history but certainly a tragedy for many of its characters)—can be well understood in rational-choice terms: their high drama is less a product of “irrational” behavior than a train of events, and rational choices in response to them, that spirals out of control.

Indeed, what start out as rather routine calculations in these works become anything but routine in their consequences for the players. Besides the aforementioned tragedies, mundane calculations go awry in an O. Henry story describing a classic coordination problem, which is compounded by incomplete information that also plagues Portia’s suitors in a game they play with her father in Shakespeare’s Merchant of Venice. But incomplete information also creates opportunities for signaling and credible commitments, which are prominent in works by Conrad, Homer, Puccini, Shakespeare, and George V. Higgins that are briefly discussed in Section 6.

In Section 7, I consider game-theoretic analyses of the devil in Goethe’s Faust and of God in the Hebrew Bible. Faust is modeled as a differential game, whereas several stories in the Hebrew Bible are viewed as simple ordinal games. The latter games are made interesting by the fact that they are interconnected by the continuing presence of God, who exhibits an abiding interest in using threats to cement His reputation and thereby tries to deter future untoward actions, including some by His “chosen people,” the Israelites.

Sir Gawain and the Green Knight, a medieval narrative poem that has been explicitly modeled as a game of incomplete information, is discussed in Section 8. In the model, reputation plays a prominent role in explaining the actions of the main characters. Also modeled is the dual character of Sir Gawain, whose two natures clash in an intrapsychic game over how to view the Green Knight.

In Section 9 I offer some observations on the “state of the art”—an apt phrase, I believe, because game theory, as applied to literature, is still more an art than a science. I also discuss new uses of the theory, such as the exploration of games played between an author and a reader that incorporate prior expectations of each player. I conclude that game theory offers an economical structure for clarifying strategic issues in plot design and character development that most literary theories do not.

2. METHOD OF INQUIRY

Besides considering the merits of different applications, it is useful to inquire how game theory has gained that foothold that it has in literary
analysis.\textsuperscript{2} For this purpose, I wrote several individuals who had applied game theory to literature and asked them the following questions:

1. What inspired you to make the application(s) you did? Are there other humanistic works you considered?

2. Does game theory offer unique insights into these works? Or does it offer more a framework for elucidating strategic conflict that these works illustrate?

3. Do these applications make a contribution to game theory, viewed as an applied field? What kind?

A number of respondents did not confine themselves just to these questions but went on the express wide-ranging views, replete with examples, of what benefits game theory can bring to the study of literature and vice versa.

To organize this rather open-ended information, I have grouped applications partly in terms of the theory (e.g., the Minimax Theorem, games of coordination) and partly in terms of literary motifs (e.g., the role of emotions, the rational foundations of tragedy). At the same time I try to give an historical perspective to the applications by reporting what influenced people, told mostly in their own words, to make the applications they did and what they see as their benefits to both literature and game theory.

Because few readers will be familiar with all the applications that have been made, I have included some information about the applications themselves, especially if they seemed representative or unique in their approaches. In one instance, I go outside the extant literature to show how William Faulkner, in a grim tale of pursuit and mayhem, better captured the unpredictability of strategies in two-person constant-sum games without a saddlepoint than the usual authors cited on this subject.

The examples I discuss illustrate how game theory can enhance one’s understanding of the strategic elements of a fictional situation. The feedback may also go in the other direction, whereby a story, for example, may force the theorist to rethink how game theory may need to be extended or refined to mirror the strategic situation that it describes.

Before discussing some of the applications and looking at responses to the questions I posed in my letter, a chronological listing of literary works to which game theory has been applied is worth perusing (see Table I).\textsuperscript{3}

Game-theoretic exegeses of these works range from a few sentences to

\textsuperscript{2} I choose the word "foothold" with care: game theory has hardly taken literary analysis by storm, perhaps in part because the theory is often misunderstood by humanists.

\textsuperscript{3} This list does not include literary works that inspired game-theoretic models but are not seriously interpreted by the model, such as the King Solomon story in Glazer and Ma (1989) and a Dostoyevsky novel, Notes from Underground, in Gilboa (1991).
TABLE 1

LITERARY WORKS TO WHICH GAME THEORY HAS BEEN APPLIED

2. William Shakespeare, *Merchant of Venice* (Williams, 1954)—play
15. Homer, *The Odyssey* (Elster, 1979)—mythology
16. Hebrew Bible (Old Testament) (Brams, 1980)—religious work

Lengthy articles. They also vary greatly in technical level, from relatively informal strategic descriptions to sophisticated mathematical analyses.

In applying game theory to literary works, it is useful to bear in mind the admonition of Howard (1971, p. 146) that “skillful authors often conceal certain essential motivations of their characters in order to reproduce the mystery we often feel in real life as to why people behave in the way they do.” Thus, game theory would seem least applicable to works of surrealism or fantasy. On the other hand, it would seem most useful in helping unravel certain mysteries—or perhaps even helping authors achieve their own ends (e.g., to surprise the reader).

3. AVOIDANCE AND ACCEPTANCE OF THE MINIMAX THEOREM

A number of conflicts in the literary works assayed can be viewed as constant-sum, in which what one player wins the other players lose. If there are only two players, the Fundamental Theorem of Game Theory,
or Minimax Theorem—proved in von Neumann (1928), 16 years before
the first edition of von Neumann and Morgenstern (1944)—establishes
that both players can choose strategies that guarantee themselves values
for the game (a positive quantity for one player and the same quantity, but
negative, for the other in a zero-sum game); however, these strategies may
be mixed.

Mixed strategies introduce an element of uncertainty into the play of a
game; that they guarantee a certain expected value, however, offers no
assurance of what will occur in any single play of the game. Consider the
scene in Conan Doyle’s *Sherlock Holmes*, in which Sherlock Holmes,
pursued by the notorious Moriarty, must decide whether to get off his
train at Dover or at Canterbury, an intermediate stop. In the story, he
chooses Canterbury, anticipating that Moriarty will take a special faster
train to Dover to try to catch him if he gets off there. Holmes’s anticipation
is correct, but Morgenstern (1935, p. 174) asks the critical question that
mixed strategies are designed to address: “But what if Moriarty had been
still more clever, had estimated Holmes’ mental abilities better and had
foreseen his actions accordingly?”

Morgenstern originally posed this question in his first book (Morgen-
 stern, 1928), which coincidentally was published the same year as the proof
of the Minimax Theorem. Unaware of the Minimax Theorem, Morgenstern
saw the Holmes–Moriarty story as an illustration of a paradox in which
“an endless chain of reciprocally conjectural reactions and counter-reac-
tions . . . can never be broken by an act of knowledge but always only
through an arbitrary act—a resolution” (Morgenstern, 1935, p. 174). While
prescient in recognizing the arbitrariness of the resolution, Morgenstern
did not yet know its mixed-strategy form, even though mixed strategies
had actually been calculated for specific games before the Minimax The-
orem was proved (Dimand and Dimand, 1990).

Conan Doyle’s resolution, on the other hand, was to make Holmes one
whit more clever than Moriarty, ignoring that Moriarty himself might
have been able to make an anticipatory calculation similar to Holmes’s.
Moreover, the matter does not end there: Holmes could have anticipated
Moriarty; Moriarty, Holmes; and so on, leading to Morgenstern’s “endless
chain” of reasoning.

In *The Purloined Letter*, Poe carried Conan Doyle’s resolution one step
further by assuming that an extremely clever boy could always calculate
exactly how far ahead less clever opponents would reason. Then, in a
game in which this boy guessed whether an opponent was concealing an
odd or an even number of marbles in his hand, the clever guesser would
be able to anticipate his opponent, whether the opponent was a “simple-
ton” or someone of great cunning (but not greater than his own). Here is
how the clever boy, according to Poe, was able to do this:
When I wish to find out how wise, or how stupid, or how good, or how wicked is any one, or what are his thoughts at the moment, I fashion the expression of my face, as accurately as possible, in accordance with the expression of his, and then wait to see what thoughts or sentiments arise in my mind or heart, as if to match or correspond with the expression. (quoted in Davis, 1970, pp. 26-27)

Labeling this reasoning "tortuous," Davis (1990) points out that "the adversary can undo all the boy's labor by simply randomizing, in which case it will take nothing short of the Delphic Oracle to gain an edge." Davis (1990) chose this example; he wrote, "because of the irony of Poe's comment: 'As poet and mathematician, he would reason well; as mere mathematician, he could not have reasoned at all'" (quoted in Davis, 1970, p. 27). On the contrary, Davis (1990) argues, "as mathematician (using the minimax theorem) he need not reason at all—random play is sufficient to confound the boy."

Hence, it is the "mathematician"—who, according to Poe, "could not have reasoned at all"—who can play this game at least to a draw, even against an incredibly clever opponent. By randomizing, the mathematician robs the opponent of any control over the outcome and so ensures the value of the game.4

This is a fundamental insight of the Minimax Theorem that neither Conan Doyle nor Poe seems to have understood. (To be sure, the cunning these writers attributed to their characters may make for better fiction than resolving each game with the flip of a coin.) But just the opponents' knowledge of this greater cunning would have been sufficient for them to even the score by choosing mixed strategies; apparently they did not have even this knowledge—or, more accurately the writers did not choose to give it to them, presumably because they were more interested in describing games of skill in which one character outwits the other.

Not all writers portray their characters in such a one-sided fashion. For example, knowledge is more shared, and calculations more even-handed, in the climactic scene of Faulkner's novel, Light in August (published in 1932), in which Percy Grimm pursues Joe Christmas, a prisoner who has just escaped his captors. Though handcuffed in front. Christmas, like Grimm, has a gun. Grimm thinks, as the pursuit by bicycle and on foot nears its end, like a game theorist: "'He can do two things. He can try for the ditch again, or he can dodge around the house until one of us gets a shot. And the ditch is on his side of the house'" (Faulkner, 1950, p. 404).

Grimm runs for the ditch, but soon he realizes that "'he had lost a point. That Christmas had been watching his legs all the time beneath the house."

4 Of course, knowing exactly how clever an opponent is, the boy can always win, but this cleverness is better characterized as omniscience, which even the biblical God did not possess (Brams, 1980).
He said, 'Good man' "(Faulkner, 1950, p. 405), acknowledging his miscalculation.

The pursuit continues until it reaches the house of Reverend Hightower, who, though knocked down and injured by Christmas when Christmas burst in, refuses to tell Grimm in which room Christmas has hidden, despite Grimm's demands: "‘Which room?’ Grimm said, shaking him. ‘Which room, old man?’” (Faulkner, 1950, p. 406). After Grimm asks once again, Hightower attempts to exonerate Christmas for the alleged murder he committed, but Grimm "flung the old man aside and ran on" (at random?) into the kitchen (Faulkner, 1950, p. 406).

A fictitious "Player"—a literary device in the novel—guides Grimm, but this guidance seems founded on no more than chance moves. As Grimm storms into the kitchen, where Christmas has overturned a table to protect himself, he unloads his revolver into the table. Before Christmas dies, Grimm castrates him with butcher knife he finds in the kitchen.

This, the most gruesome scene in the novel, contrasts sharply with Grimm's pursuit of Christmas, which is all cool calculation. Faulkner seems to have invented Player to epitomize the calm and deliberate mind of the fanatic; Grimm, who is "moved" by Player, is utterly devoid of any emotion, except when he explodes with savagery in the end. The beast in Grimm coexists with the cerebral Player, which is a juxtaposition that game theory normally does not entertain when it posits a player with one set of preferences.5

Unlike Conan Doyle and Poe, Faulkner beautifully captures the uncertainty inherent in mixed strategies. In the face of the uncertainty created by an inscrutable opponent, players still must act. And act Grimm does: first to his own disadvantage when he discovers that Christmas could follow his movements as he ran toward the ditch; second to his advantage when, "waiting for Player to move him again" (Faulkner, 1950, p. 406)—presumably in some randomized fashion—he rushes the kitchen. Faulkner has little to say about the motivations behind Christmas's choices, but it seems they were also essentially arbitrary.

Faulkner does not assume that one player had superior computational abilities. True, Grimm has Player on his side, so to speak, but this device, in my view, reinforces the desultory character of Grimm's choices. Calculated they may have been, but because Grimm, at each stage of the pursuit, had only imperfect information, he could never be sure what his best choice was. Grimm "won," finally, not because of sheer cleverness but because the game was unfair—the odds were heavily stacked against the fugitive, Christmas, whom Grimm so relentlessly hunted down.

5 If more than one type of player is allowed, as in games of incomplete information, only one type is actually the true type—there are not different types embodied in a single player (e.g., with multiple personalities).
I have offered this analysis of a scene from *Light in August* to show that Faulkner is one fiction writer who had an astute if implicit understanding of mixed strategies in two-person constant-sum games of imperfect information. Doubtless, other examples could be found. While the scenes that Morgenstern and Davis discussed in *Sherlock Holmes* and *The Purloined Letter* have the earmarks of games in which mixed strategies are optimal, both Conan Doyle and Poe shrank from making their protagonists' opponents as smart as the protagonists themselves. They got tidy results that way, but the minimax solution in games of imperfect information shows that not all conflicts can be resolved by outguessing. Faulkner understood this.

4. Are Constant-Sum Games Emotionless?

Constant-sum games with more than two players raise entirely new theoretical questions, chiefly related to what coalitions are likely to form and remain stable. Howard (1990) reports that when he went to a performance of *The Caretaker*, he was struck by its similarity to the game of Split the Dollar—where a dollar (or better 99¢) is divided equally among three people unless at least two agree on another way of dividing it. As you know, . . . however the dollar is divided, there are always two players who can do better by agreeing on another split which excludes the third. In the case of *The Caretaker*, there is a pecking order of respect, such that the least-respected character can always suggest to one of the other two a deal in which they give each other greater mutual respect at the expense of the third. Each of the three acts deals with the formation of one of the three two-person coalitions.

Howard (1971) describes the formation and disintegration of each coalition in the three acts, involving two brothers who share a house and a third man who might become their caretaker. The play ends with "no relationships," but with the possibility that new relationships will form once again, "causing the three acts to be repeated in sequence again and again" (Howard, 1971, p. 145). Although *The Caretaker* "is almost classically austere and simple from a game-theoretic point of view" (Howard, 1971, p. 145), Howard argues that "Pinter's view is however interesting in that at least he has risen to the level of dramatizing a three-player interaction" (Howard, 1990).

In analyzing C. P. Snow's *The Masters*, Riker (1986) examines the more complex interactions of 13 fellows in a Cambridge college, who must vote on a new master of their college in a constant-sum game (there are two candidates, and only one can win). The novel is about the election campaign, in which "pride and ambition and humiliation and failure are displayed against a background of political bitterness" (Riker, 1986, p. 52).
There are four switches in support for the two candidates as they vie for the votes of the other 11 fellows of the college. Riker shows how the maneuvers of one fellow, in particular, who abandons his original favorite the day before the election, ultimately succeed. Although Riker's analysis stresses social-choice theory rather than game theory (e.g., Riker shows that no logrolling is possible, based on the positions of the fellows on two dimensions), it is evident that the campaign is suffused with game-theoretic calculations.

The leaders of the two factions constantly plot to hold their coalitions together, and draw in new members, against opposition efforts to woo away potential defectors. Riker in fact explored this idea in an earlier game-theoretic model (Riker, 1962); its best-known prediction—that only minimal winning coalitions will form under certain assumptions (the so-called size principle)—is exactly what happens in Snow's story.

Riker regards *The Masters* as uniquely political: it "is, so far as I know, the only one [novel] in which politics is not mere background but the very plot itself" (Riker, 1986, p. 52); "all other novels concern character development, love affairs, hurried journeys, family history, etc."

Riker admits that building coalitions is "hardly the stuff to release readers' adrenaline as do seductions, quarrels, or chases," but he believes "political ambition, and indeed political success, uniquely reveal tragic flaws in character," as demonstrated by Greek dramatists and Shakespeare (Riker, 1986, p. 52).

To Howard (1990), by contrast, most interesting conflicts are not zero-sum:

Such a zero-sum view is a common one, as shown by the frequent comparisons of politics or war with chess, poker or football. I think it is unrealistic: all my experience with applying game theory leads me to think that people are both more clever than this (they don't see things as zero-sum when they aren't) and more stupid (the simplest game-theoretic model of their situation often shows them simple, gross obvious things they have entirely failed to see).

For Howard, "Pinter's view is the bleak, cynical one obtained by supposing that adults do not grow out of the 'zero-sum' mentality of children." He disdains this view, adding that it is "a rare, sophisticated aberration of the 20th century elitism...and means [the] absence of emotion, deceit, preference change, etc.—all the things that artists have traditionally been most interested in" (Howard, 1990).

In Section 5, I turn to applications of game theory that have been made to nonconstant-sum games. Whether game theory can illuminate emotions in such games, as Howard maintains, or whether it is better suited to elucidating purely political plots and stories, as Riker maintains, is a question whose answer may shed light on the type of literature that has been selected for game-theoretic scrutiny.
In depicting $n$-person games of coalition formation, Pinter and Snow illustrated the fragility of coalitions in constant-sum games, which Howard (1971) and Riker (1986) explicated by showing how alliances may unravel. From a literary point of view, however, the question is not the stability of coalitions but whether such works are all plot and calculation—or something more. And does the "something more" require that characters transcend their own rationality?

5. The Rationality of Tragedy

The early use of game theory in literary exegesis includes Rapoport's interpretations of Othello and Tosca as nonconstant-sum games (Rapoport, 1960, 1962). In a two-person normal-form version of Othello, Othello may believe or not believe that Desdemona has been faithful, and Desdemona may deny or confess (falsely) her guilt; the tragedy occurs when Desdemona denies that she has given herself to Cassio, but Othello, with seeds of doubt planted by Iago, does not believe her.

Rapoport also considers an extensive-form version of this conflict, involving the four principals and "chance," which has 55 distinct outcomes. The enormity of Desdemona's 16,384 strategies in the game tree leads Rapoport to remark that "perhaps enough has been said about the practical difficulties of applying game theory in human affairs" (Rapoport, 1960, p. 240). But he argues that "game theory stimulates us to think about conflict in a novel way" (Rapoport, 1960, p. 242; italics in original) and also shows how interdependent decision situations can be "precisely characterized and rigorously analyzed" (Rapoport, 1990).

After analyzing Othello, Rapoport turned to Tosca (Rapoport, 1962), which he analyzed as a $2 \times 2$ Prisoners' Dilemma. Whereas jealousy fuels the plot in Othello, it is Tosca and Scarpia's mutual betrayal in Tosca that leads to its tragic end.

In Teodorescu-Brinzeu's (1977) analysis of Othello, she assumes that Othello and Iago are involved in a constant-sum game, which, especially from the perspective of Othello (who is sympathetic to Iago until the end), seems to me a misinterpretation. Second, she assigns payoffs so that Iago has a dominant strategy, and Othello a best response, but then argues that this "wise [minimax] solution" was not chosen because "it lacks dramatic consistency as it is very commonplace." Instead, "the psychological reality requires that in this clash of passions the Moor's jealousy and Iago's hatred should overcome any lucid calculations and drive them both to destruction"; indeed, they "die devoured by their own passion" (Teodorescu-Brinzeu, 1977, p. 373).

Coupled with Desdemona's murder—not to mention Roderigo's and
Emilia's—this tragedy suggests to me that there were no winners, making the game decidedly nonconstant-sum. Thus, I think the interpretation of Othello as a constant-sum game is untenable.

Teodorescu-Brinzeu's (1977) contention that great drama may require that the characters reach beyond themselves (irrationally?) to seize the moment—sealing their fate and, quite often, their destruction—deserves further comment. This view seems to be a tenet of Marcus's (1977) "Romanian School of Mathematical Linguistics and Poetics," because it is also reflected in Lalu's (1977) game-theoretic analysis of Shakespeare's Richard III.6

Lalu (1977) analyzes this play as an extensive-form nonconstant-sum game and concludes that

what the playwright considers as the optimal strategies are in fact optimal for the tension and the rhythm of the performance, seldom for the "actual life" of the character. A cautious hero would be uninteresting. Paradoxically, the optimal strategy of the character is, more often than not, that of "the mad risk." Therefore, the main characters may seldom be considered as perfectly rational players; as far as we view the play in terms of "a slice of life," the characters make mistakes. The optimal strategies for their destinies of actual human beings will seldom be followed; on the contrary, the characters will act following those strategies which the author (perhaps the only rational player) thinks optimal according to an aesthetic criterion. (Lalu, 1977, p. 343)

Lalu (1977, p. 343) then asks what the point of applying game theory is and answers that she is interested in exploring deviations from rationality that are "optimal within the frame of the whole play, regarded as a work of art."

In my opinion, there is considerable arbitrariness in Lalu's (1977) assignment of specific numerical values to outcomes, and specific probabilities to chance events, in Richard III. These assignments vitiate her claim that Richard chose his worst strategy, though I would not dispute her claim that Shakespeare sought "the ruin of the character [Richard] . . . for the sake of the tension of the performance" (Lalu, 1977, p. 349).

The issue is whether this tension was achieved by making Richard's choices irrational. I think it was not, and an alternative and more defensible game-theoretic analysis—not to be developed here—could, I believe,

6 On the other hand, Steriadi-Bogdan (1977), also a disciple of this school, argues in a game-theoretic analysis of Christie's The Mousetrap that the characters made rational choices. But The Mousetrap is a detective play that is not generally considered to be a great tragedy, whereas "in studying Shakespeare's Othello, namely Iago's strategy, you have to observe that Iago does not look for what in the Mathematical Game Theory is called the best strategy, but rather for the worst strategy" (Marcus, 1990; italics in original). I remain unconvinced that Iago chose, say, a dominated strategy—at least in the beginning, when his plan seemed to be working quite nicely—but I agree that combining "strategic and psychological aspects . . . is a rather delicate task" (Marcus, 1990).
demonstrate that Richard was eminently rational. Briefly, the argument underlying this alternative interpretation is that Richard, brilliant and diabolical, knew that he could act boldly with a high likelihood of success; in fact, he rapidly dispatches several of his opponents at the beginning of the play. Although lacking the contemplative character of a Richard II or Hamlet, who seem to weigh options more carefully, as Lalu points out, Richard III, nevertheless, seems no less rational (and tragic) a hero.\(^7\)

I agree with Lalu that Richard III is not prudent, but prudence, which Lalu equates with the minimax principle and estimating lotteries, is not synonymous with rationality. And neither is a hero’s “tragic fall” synonymous with irrationality. In fact, contrary to Lalu, I believe the tragic fall is made more, not less, poignant when characters are driven by an inexorable rationality toward some terrible end.

6. Coordination Problems, Signaling, and Commitment

Unlike Othello and Tosca, in which the characters displayed a stunning lack of trust in each other—for good reason in Tosca but less so in Othello—the theorists who have analyzed O. Henry’s The Gift of the Magi see the conflict as arising for almost the opposite reasons. The husband, who sells his watch to buy his wife combs, and the wife, who sells her hair to buy her husband a watch fob, are blinded by their love and perhaps too trusting. Their blindness leads to a failure to coordinate their gift giving, and great sadness in the end (at least for the reader—more on the game between the author and the reader later) when the consequences of each’s trying to surprise the other are discovered. Rapoport (1960, p. 171) speaks of the couple’s “misplaced altruism.” Vorob’ev (1968, pp. 370–372) views the game as a Battle of the Sexes (the usual story of this classic game is given in Luce and Raiffa, 1957, pp. 90–94), and Rasmusen (1989, p. 40) argues that the couple’s failure to communicate may, ironically, have been rational because communication would have ruined the surprise. Indeed, their sacrifices affirmed their great love for each other, despite their misfortune.

Rasmusen (1990) points out that the couple, in effect, chose a mixed-strategy equilibrium;\(^8\) the pure-strategy equilibria would be the outcome in which either the husband or the wife gives a gift but the other does

---

\(^7\) Incidentally, the historical Richard III seems to fit this portrait well (Ross, 1981).

\(^8\) Williams (1954, pp. 201–203) discussed such an equilibrium as the solution to a “marriage game” in Shakespeare’s Merchant of Venice. But this game, which is between Portia’s father and her suitors, is constant-sum, because the father wants to frustrate, not coordinate with, Portia’s suitors.
not. Although game theory tells us that the mixed-strategy equilibrium is inefficient, and may be disastrous when the players choose noncomplementary mixed strategies (as occurred in the story), it does not tell us how such a dismal state of affairs may arise. By contrast, the story suggests that

the act of communication would lower utility by eliminating the fun of being surprised. So the example says something about how to apply the theory. The theory also says something about the example: that even if the two people suspected that the ridiculous outcome might occur, they might do it anyway. And it also makes you think about what might have been one of O. Henry's points, that it is the thought that counts in giftgiving. (Rasmussen, 1990)

Rasmussen (1990) draws a larger lesson from such material:

In general, examples are good for suggesting wrinkles that might not otherwise occur to the theorist. The easiest way to break out of a paradigm is to have the real world suggest a problem with it. Since often the scholars are too used to thinking in one particular way. It is perhaps harder to be surprised by theories than by data.

I concur with these views but do not know of any direct evidence whereby a game-theoretic analysis of fiction has generated significant new theory. On the other hand, a large literature on "signaling games" that has developed in recent years is germane to the strategic exegesis of plots. Dixit (1990) gives a contemporary example:

If you read past all the four-letter words and the graphic violence, the whole theme of Cogan's Trade by George V. Higgins [1985] is reputation. For reasons too complicated to explain in brief, the bosses of organized crime in Boston have lost their reputation for protecting the activities they sponsored. How to regain it? This is a signalling game, and as usual there is excessive investment in signalling, in this case quite literally overkill. And the theory of this is almost fully and correctly explained by the enforcer (Cogan) in a conversation with The Man's counsellor.

Schelling (1960, p. 140; 1966, pp. 11, 37) offers examples of the subtle and not-so-subtle signaling of threats in Shakespeare's Henry V and Measure for Measure and Joseph Conrad's The Secret Agent. Why the threat of punishment for refusing to conform to the demands of Communism is

---

9 Why these literary choices? Schelling (1991) reports

that my use of Henry V in Arms and Influence came from just seeing the play in London in 1965; I certainly didn't go to the play looking for illustrative material. I have no recollection of Measure for Measure, but I must have seen it in New Haven on the stage because I cannot imagine that I ever would have read it. . . . I do specifically remember how I was led to Conrad's The Secret Agent. I heard it from Daniel Ellsberg and when I wanted to use it I called him up and asked whether he was planning to use it in print in the near future and he said no and I asked whether he would release it and he said yes and I read the book and found no other useful examples but did use that one.
omnipresent in *Dr. Zhivago* (Howard, 1988). Citing different passages from *Henry V*, Dixit and Nalebuff (1991, pp. 161–162) show how Henry inspired his troops, and thereby made his commitment credible, before the battle of Agincourt. His "steel my soldiers' hearts" prayer echoes Lady Macbeth's plea to the "spirits," as she plans the murder of King Duncan in *Macbeth*, to "Make thick my blood/Stop up access and passage to remorse/That no compunctious visitings of nature/Shake my fell spirit."

Going one step further, in Homer's *Odyssey* Ulysses has himself bound to ensure that he will not succumb to the temptation of the sirens (Elster, 1979, p. 36).

To be sure, the aforementioned examples of threats in works of fiction should not be construed as game-theoretic models. Harper (1991) goes a step further in analyzing the credibility of threats in Puccini's opera, *Gianni Schicchi*, which involves deception in the awarding of an inheritance. He shows how different interpretations of the game lead to different extensive-form and normal-form representations and how various refinements single out different Nash equilibria.

7. THE DEVIL AND GOD

In Goethe's *Faust*, Faust gambled not just his wealth or reputation but also his life in making a compact with the devil. By selling his soul to Mephisto in exchange for knowledge and power for 24 years (in other versions of the Faust legend, sex or youth is the lure), Faust appeared to commit himself irrevocably to eternal damnation when the "supreme moment" arrives. Fortunately for Faust, his final repentance saves his immortal soul from Mephisto, though not all versions of this legend have such a felicitous ending.

Mehlmann (1988, 1989) uses differential game theory to analyze Goethe's great drama, making certain assumptions about the linkage between the players' beliefs about the timing of the supreme moment and also about how the players' payoffs are affected by each other's activities (repentance by Faust, temptation by Mephisto). He demonstrates consequences of these assumptions for the equilibrium path, arguing that Faust's "will to strive" (i.e., to repent), as the supreme moment approaches, explains his salvation.

Mehlmann (1990) cites other purported explanations (literary, legal) for Faust's salvation but claims that his mathematical model has "all the ingredients needed." Although I am not convinced that he has captured the essence of the drama in the parameters and functions he assumes, his application illustrates how advanced tools of game theory can be employed in literary exegesis.
Mehlmann (1990) reports that he has "always been interested in unusual applications of mathematics" and believes that "mathematics should play the role of an art rather than that of a science." Searching for a dynamic conflict situation to which he could apply differential game theory, "by chance . . . Faust came into my mind." He says that this application makes contributions both to the mathematical theory and to the modeling of player beliefs.

My motivation for applying game theory to the Hebrew Bible came from teaching a so-called humanities seminar at New York University for freshmen and sophomores, in which one requirement was that "primary sources" be used. I hoped to show how, through a careful reading of certain narratives in the Hebrew Bible, elementary game theory could lend coherence to the strategic interpretation of these stories. I also hoped that the analysis of several individual stories would allow me to draw general conclusions about the "games" biblical characters (God included) played. The seminar, which included orthodox Jews, devout Catholics, fundamentalist Protestants (and others) turned out to be very stimulating and led to Brams (1980).

In this book, I show that biblical characters are by and large rational in the 20 or so stories of conflict and intrigue that I analyze. God is a "superlative strategist," but having granted free will to His human subjects (which justifies a game-theoretic treatment), He is besieged by problems that their freedom engenders. These cause Him great anguish, leading to very human-like displays of anger, frustration, and jealousy.

God's wrath is especially great when his chosen people, the Israelites, cross Him. It is sometimes expressed in petty, sometimes vindictive, behavior, but He is also merciful, always stopping short of wiping the slate clean.

I believe that the Bible, as well as other religious works regarded as sacred, can be viewed at two levels. One level is as a literary work, with the stories it tells being susceptible to the same kind of game-theoretic analysis that helps to make the strategic aspects of secular stories perspicuous. The other level takes account of religious questions, such as the rationality of belief in a superior being or the problem of evil, which I have addressed in another work (Brams, 1983).

The profound and profane may not be so different, at least in terms of the kinds of game-theoretic models needed to explicate their strategic structures. For example, if a superior being is immortal, this being must be concerned with its reputation, which in fact obsesses the biblical God, especially in the Torah (the first five books of the Hebrew Bible). Thus, it makes sense to consider a concern with reputation as a correlate of immortality, to which the substantial recent literature on reputation in repeated games is pertinent.
8. Reputations and Intrapsychic Games

Reputations are based on beliefs, and the modeling of such beliefs is a central feature of O’Neill’s (1991) application of game theory to Sir Gawain and the Green Knight, a Middle English poem of the late 14th century (and only rediscovered in the 19th century). This poem describes the sudden appearance of a Green Knight of immense size who challenges the hero, Sir Gawain, to behead him in exchange for a return blow. After accepting the dare, which results in the beheading but not the death of the Green Knight, the poem recounts Gawain’s search for the Green Knight, including tests of chivalry he must endure, before the Green Knight is allowed his turn to behead Sir Gawain a year and a day later. Feigning a beheading, the Green Knight inflicts a minor wound on Sir Gawain, presumably to symbolize Gawain’s imperfection.

Now considered a great literary work that is rivaled only by Chaucer’s poetry of the same period, Sir Gawain “engages modern readers by addressing modern problems,” in particular “the predicament of how to follow one’s ideals when the world maneuvers them into opposition to each other” (O’Neill, 1990). Although the story might seem fantastic, the Green Knight is not described just in supernatural terms but is given a distinct human dimension, suggesting him to be vulnerable emotionally if not physically.

O’Neill (1991) analyzes two games, the first having to do with Gawain’s reputation, which the Green Knight throws into doubt by his bizarre challenge to Arthur and the Round Table (Gawain persuades Arthur to let him stand in for him). O’Neill (1991), postulating different beliefs that the players might have in different versions of a game of incomplete information, analyzes why the Green Knight throws down the gauntlet and why Gawain accepts.

In one version, for example, he argues that Gawain seeks to enhance his reputation for placing a high value on his reputation, defined recursively. In other words, Gawain wants to be seen as someone to be reckoned with generally, independent of the specific challenge he faces. The Green Knight makes a similar calculation in uttering his dare, and the players compete in a contest to bolster their relative reputations.

The second game O’Neill (1991) analyzes is that between Gawain’s two natures—one chivalrous and the other self-preserving—that recalls the conflict between the id and the superego in Freud’s theory. (The third component in Freud’s theory, the ego, might come into play if there were

---

10 O’Neill (1991) also analyzes The Feast of Bricriu, an Old Irish medieval tale that describes another beheading, but its analysis is similar to that of Sir Gawain, so I shall not discuss it here.
a mediator or arbitrator involved.) In effect, Gawain must play against himself, not knowing whether the Green Knight is (1) chivalrous and vulnerable or (2) malevolent and invulnerable, which would make the game fair or unfair, respectively. If (1), then Gawain owes the Green Knight fair play, which will be reciprocated; if (2), then Gawain is absolved of his duty to rise to the challenge and should instead avoid being killed.

There is psychic harmony in this game if Gawain's two natures agree on the character of the Green Knight, but each of the natures prefers a different interpretation: the chivalrous nature prefers (1), the self-preserving nature prefers (2). If the two natures disagree, there is tension, which is worse for both players (i.e., Gawain's two natures) than harmony. The resulting game, in which the two natures are locked in battle, is Battle of the Sexes, which has two Nash equilibria in pure strategies and one in mixed strategies.

The lack of an obvious solution, O'Neill (1991) argues, renders the outcome equivocal, which "makes for a good literary plot." Unlike the Romanian school, however, O'Neill (1991) does not contend that a character must act irrationally in order to dramatize the conflict. Instead, the players' harrowing choices, due largely to the coordination problem caused by the lack of information on how to regard the Green Knight, sustain our keen interest in the story.  

Which, if either, persona of Gawain has its preferred outcome chosen (the chivalrous nature prefers a chivalrous Green Knight, the self-preserving nature a malevolent Green Knight) depends on how the intrapsychic battle between Sir Gawain's two natures is resolved. The actual resolution in favor of chivalry validates Sir Gawain's acceptance of the dare, but through most of the narrative the rationality of this course of action is anything but apparent.

O'Neill's motivation for analyzing a literary work is very different from Mehlmann's:

I wasn't looking for a place to apply game theory. Instead I was reading the work . . . and then it occurred to me that it was an interesting problem to formulate the hero's situation as a game . . . Some of the hero's problems were the same as problems in my life at the time . . . , and this led me to think very hard about the poem. I read it and pondered on it. I would walk around thinking about it. It was not just for entertainment (O'Neill, 1990).

Like other theorists, O'Neill (1990) believes that game theory can clarify a literary work. Nonetheless, he points out that some "past applications

11 A reader, in my view, is much more likely to identify with a rational protagonist than an irrational one, especially one, like the Green Knight, who seems so unbelievable from the start.
of game theory . . . did not take the literary work seriously in its details’’ nor take account of ‘’good ideas scattered through the informal literature.’’

9. Future Prospects

Besides taking the textual details of a literary work seriously, O’Neill (1990) claims that ‘’it is also necessary to relate our work to the vocabulary already in use’’ if game theory is to make a contribution to literary analysis. (He is less sanguine that the game-theoretic analysis of literature will lead to mathematical advances.) More practically, O’Neill (1990) is concerned that neither literary nor mathematics journals are generally open to linkages between these very different disciplines.

It is difficult to say how much the lack of publishing outlets has retarded interdisciplinary work. My own belief is that linkages between mathematics and literature are not viewed as worth exploring by young scholars in either field if they are interested in advancing their careers. aggravating this problem is that there is no interdisciplinaty training for people who might be interested in the combination, with the possible exception of the Romanian school mentioned in Section 5.

A related problem is that several of the applications I have discussed are no more than off-the-cuff illustrations. While most of the authors are mathematically sophisticated, they seem to have made little effort to find nontrivial applications of game theory. Of course, they cannot be faulted if a probing literary analysis was not their objective, but still one might hope for a more serious concern with the literary work. O’Neill (1990) speculates that Vorob’ev, a respected Russian mathematician who offers cursory analyses of several fictional works (see Table 1) but did not report his own views, ‘’perhaps regarded his study of game theory and literature as an interesting diversion, reading for the masses.’’ People like Vorob’ev with the technical skill to do serious game-theoretic analysis are not attuned, it seems, to the more subtle literary issues that might be modeled.

On the other hand, in cases in which the literary work is primary, the game-theoretic analysis is sometimes flawed (true of some of the Romanian authors). The opposite problem plagues Mehlmann (1988, 1989), wherein the mathematical structure is impressive but is not persuasively related to the narrative.

Both Howard (1971) and O’Neill (1991) use nontrivial game theory to construct plausible strategic interpretations of the works they analyze. Interestingly enough, both authors, as noted earlier, indicated that they did not set out to ‘’apply’’ game theory, but the literary works themselves riveted their attention.

Other tools of mathematical analysis have been applied to literature,
but they generally give short shrift to plot (some citations are given in O’Neill, 1991). Game theory makes plot front and center: when there is no strong plot or story line, as is the case in much modern fiction, then the theory probably has little to offer. I share Howard’s (1990) view that “plot is essential for the kind of great art which really changes people.” so I am not worried that game theory will suffer from lack of good literary material to which to apply its methods, some modern fiction notwithstanding.

Howard (1990), who reports that he analyzed “every incident and conversation as a set of interlinked games” in Anthony Trollope’s *The Warden* and then transposed the novel into a modern setting (Howard’s reworked version was not published), indicates that game theory may have other roles to play, such as

> to help writers construct plots. In film-making, where many people have to cooperate, it would be exceedingly useful to work with a clear game-theoretically analyzed plot—just as musicians find it useful to have a score.

He adds that this kind of analysis can also help game theory, because game theorists

> benefit from the great store of intuitive wisdom about human behaviour contained in the world’s fiction. They should continually be testing their theories against this.

> If it doesn’t make sense to Shakespeare, perhaps it doesn’t make sense!

Game theory, in my view, should be able to do more than suggest that there is a problem in a relationship. The fact that Scarpia and Tosca are enmeshed in a Prisoners’ Dilemma, or that the husband and wife in *The Gift of the Magi* have a coordination problem à la Battle of the Sexes, is not particularly enlightening. Why are these stories compelling and not just humdrum illustrations of these games?

O’Neill (1990) suggests that the tragic or surprising aspects of these stories require that we look more deeply into the information available to the players, and how it is used, in order to understand their human dramas. Indeed, the lack of information may itself be a central strategic feature of a story, as I tried to show by the players’ choices of mixed strategies in *Light in August*.

The game played between the author and the reader, as the reader progressively acquires more information (not necessarily accurate, e.g., the false clues in a mystery), is one that does not seem to have been analyzed for any literary work. An appropriate framework for such an

---

12 However, this subject is the main theme of a novel, later made into a movie, by Stephen King (1987), in which a reader takes revenge on an author for killing off her favorite character in the last of a series of novels, forcing the author to burn the manuscript of his next novel and resurrect this character in a new novel.
analysis might be the theory of psychological games (Geanakoplos et al., 1989) and information dependent games (Gilboa and Schmeidler, 1988), in which players' payoffs depend on whether certain postulated beliefs are fulfilled. Thus, a reader may be either thrilled or disappointed not only by the way a story evolves but also by whether he or she is surprised by the ending. For example, an horrific ending that turns out only to have been a dream may make the reader feel either manipulated by the author or relieved, depending on the reader's prior expectations.13

I conclude that game theory and literature have their own coordination problem, with game theorists and literary analysts not often benefiting from each others' insights. What makes a literary creation is not just its overall structure but its details, including the emotional lives of its characters. Game theorists need to ponder these and adapt their theory accordingly, just as literary scholars need to appreciate that game theory has its own richness that goes beyond mathematical symbols and abstract forms.

REFERENCES


13 We sometimes use the term "cop-out" when we feel betrayed by the author. To avoid this feeling, authors might try to take account of the expectations of their readers in constructing plots and portraying characters. Thus, mystery writers might aim to surprise their readers, whereas other writers might prefer no surprises in order to stress the unrelieved boredom of the human condition (think of Samuel Beckett's Waiting for Godot). If game theory were used to help authors in this manner, it would radically change the theory's purpose: instead of using it to show that characters in a text act rationally, one would start with the characters' (or the author's) motives and write the text to show the rational working out of these motives, reversing the order in which game theory is applied from after the text is written to before.