William Godwin and Edith Wharton, Decision Theorists

John A. Weymark

Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA.
E-mail: john.weymark@vanderbilt.edu

November 2022

Abstract. Literary works by William Godwin and Edith Wharton are shown to anticipate some fundamental features of modern decision theory. In her 1904 short story, "The Last Asset," one of Wharton's characters advocates a rudimentary version of the idea of choosing a strategy in advance that specifies future choices in a dynamic decision problem. In commenting on the method employed when he composed his 1794 novel, *Caleb Williams*, Godwin said that he developed the plot and prepared outlines of its three volumes in reverse order, thereby articulating an informal version of backwards induction.

Journal of Economic Literature classification numbers. B12; B31; C73.

Keywords. Backwards induction; forward-looking decision-making; history of decision theory; William Godwin; Edith Wharton; Caleb Williams.

---

1. Introduction

The main thesis in Michael Chwe's book, *Jane Austen, Game Theorist*, is that "Austen systematically explored the core ideas of game theory in her six novels, roughly two hundred years ago" (Chwe, 2013, p. 1) and, therefore, she should be considered as one of game theory's pioneers. Underlying Chwe's argument is that contributions to game theory need not be mathematical; they can also be made using prose. I want to make a similar case for William Godwin and Edith Wharton as decision theorists.

Rational choice theory is an approach to decision-making that regards an individual as making choices so as to obtain his most preferred alternative from among those that are feasible given his beliefs about the relevant factors that are outside his control. Someone is regarded as being rational to the extent that his behavior can be explained in terms of maximizing consistent preferences given the constraints that he is facing and the beliefs that he holds. In addition to its use in the analysis of individual decision-making, this concept of rationality plays a fundamental role in game theory, which studies decision-making in situations in which there is strategic interaction. In such situations, the outcome that an individual obtains depends not only on his own decisions and his beliefs about exogenous events but also on what he believes about the decisions and beliefs of the other individuals that he is interacting with.

When choices are made over time, in order to best pursue his objectives, a decision-maker needs to anticipate what choices will be made in the future whenever he is called upon to act. That is, he should engage in *forward-looking contingent decision-making*. This is the case both when there is a single decision-maker (an individual sequential decision problem) and when there are many decision-makers (an extensive form game). *Backwards induction* is a reasoning procedure in which an optimal sequence of individual decisions or a subgame perfect Nash equilibrium in a game is determined by iteratively reasoning backwards in time.

In her 1904 short story, "The Last Asset" (Wharton, 1904), one of Edith Wharton's characters recommends planning in advance what to do in the future when called upon to make sequential choices. In advocating for the adoption of a forward-looking approach to decision-making, Wharton has enunciated some, but not all, of the features of the modern idea of choosing a strategy in advance that specifies future choices in a dynamic decision problem.

It was not until 1950 that Leonard Savage (Savage, 1972, p. 16) provided a formal model of forward-looking contingent decision-making. Planning everything in advance is what Savage calls a *grand-world problem* (p. 84). Savage contrasts this way of decision-making with that of deferring decisions until the times that they need to be made. He regards these two approaches to decision-making as capturing the intuitions underlying the aphorisms, "look before you leap" and "cross each bridge when you get to it" (p. 16). Savage acknowledges that making a complete contingent plan is not humanly possible but suggests that contingent planning is tractable if the problem one is faced with is sufficiently simple and can be considered in isolation, what he calls a *small world problem* (pp. 16–17).[1]

In the Preface to the 1832 edition of his novel, *Fleetwood* (Godwin, 1832), Godwin describes the method he employed when he composed his 1794 novel, *Caleb Williams* (Godwin, 1794). *Caleb Williams* was published in three volumes. According to Godwin's later account, he developed the plot of his novel in reverse order; first the third volume, then the second, and finally the first. Furthermore, he prepared outlines of the three volumes in the same reverse order before turning to the composition of his novel in the conventional order. Although stated informally, Godwin articulated a version of the method of reasoning that we now know as backwards induction.

The first documented use of backwards induction in a formal individual decision problem was made in 1875 by the mathematician Arthur Cayley (Cayley, 1875). Cayley considered an individual who sequentially draws tickets and receives the amount written on the last one drawn. Knowing the number of tickets, their values, and the number of draws that are allowed, the question is: When to stop drawing? Cayley's problem is an example of an optimal stopping problem—deciding when to stop sampling a random process in order to maximize some payoff.[2] It was not until the 1940s that backwards induction as a procedure for solving individual sequential decision problems came into its own, reaching its apogee with Richard Bellman's work on dynamic programming (Bellman, 1959).[3] The first explicit use of backwards induction in game theory was by von Neumann and Morgenstern (1944, Chap. III, Sec. 15).[4]

I do not claim that the informal accounts of sequential decision-making in the face of uncertainty and of backwards reasoning offered by Wharton and Godwin fully capture all that is now understood by the concepts of forward-looking contingent decision-making and backwards induction. Rather, I argue that both authors have identified what are recognizably some of their

fundamental features. Both Wharton's narrative and Godwin's account of his method of composition are silent about some of the details of the decision-making procedures that they are concerned with. As a consequence, a modern decision theorist would regard their descriptions as being not fully specified.

Literary and religious texts have been a rich source of material for illustrating game theory's concepts and methods of analysis. Game theory has also been used to interpret and critique such texts, thereby offering a nontraditional methodology for literary criticism.[5] As part of the creative process, an author may take account of the likely responses of his audience to what they read. Readers, in turn, may form expectations about what effects the author was intending to create. So, game theory is sometimes also relevant to the reading and writing of a literary text.[6] Furthermore, game-theoretic literary studies may be illuminating when some of the characters do not conform to the rationality principles employed by the theory. Indeed, a character's departure from rationality as understood by game theorists may be an essential feature of a story. For example, this is the case with folktales that feature tricksters who take advantage of the lack of strategic sophistication of their adversaries.[7] Considering these departures from standard rationality postulates raises the question of how this behavior can be formally modeled, thereby providing a source of open problems for game theorists.[8]

The relations between the rational choice theory of individual decision-making and literature have not received the attention that has been given to the relations between game theory and literature. Two scholars who have explored the connections between rational agency and literature in some detail are Jon Elster and Paisley Livingston.[9] Both Elster and Livingston provide a number of examples in which the behavior of fictional characters can be best understood in terms of purposeful rational decision-making. Elster (2009, p. 5) contends that "[r]ational choice explanation plays a critical role in the interpretation of both texts and social behavior. The interpretation of fiction benefits from the use of this type of explanation." Elster is also sensitive to the limitations of this kind of explanation, noting the importance of emotions and other non-rational phenomena in decision-making. Livingston (1991, p. 52, emphasis in the original) argues "that there are *no* literary phenomena that can be adequately understood or explained without relying, at least implicitly, on a rationality heuristic." Livingston does not make the stronger claim that it is the conception of rationality employed in rational choice theory that has this feature. Rather, his claim is about theories of

rational agency in general. Indeed, Livingston (1991, p. 51) goes so far as to say that "[i]n my view it seems safe to assume that the canons of Bayesian decision theory are not applicable to the kinds of choices, preferences, and actions that may be associated with literature ..." My contention that Godwin and Wharton are early decision theorists belies this claim. Particularly pertinent for my analysis of Godwin's 1832 Preface is that both Elster and Livingston maintain that by applying concepts of rationality to a literary text, one's understanding of the choices made by an author in composing a work of fiction can be enhanced (in addition to insight about the choices made by the characters in the narrative).

The plan of the rest of this article is as follows. Some biographical information about Godwin and Wharton is provided in Section 2. Section 3 describes how an individual sequential decision-making problem, forward-looking contingent planning, and backwards induction are formally modeled. Wharton's narrative account of forward-looking decision-making is considered in Section 4, followed in Section 5 by a discussion of Godwin's use of backwards induction as a methodology for literary composition. Some brief concluding remarks are offered in Section 6.

2. Godwin and Wharton

William Godwin (1756–1836) was a philosopher, novelist, journalist, historian, and publisher.[10] His immediate family included some of the leading literary and intellectual figures of his time: his wife, Mary Wollstonecraft; his daughter, Mary Shelley; and his son-in-law, Percy Bysshe Shelley. Godwin was raised as a strict Calvinist and briefly served as a minister, resigning from the clergy in 1782 for ideological reasons in order to take up writing. His *An Enquiry Concerning Political Justice* (Godwin, 1793) is a seminal treatise on philosophical anarchism. This book reflects the enthusiasm for overthrowing authoritarian regimes in the early days following the French Revolution. Central to Godwin's philosophy is the republican conception of liberty as independence from arbitrary power whatever its origins, governmental, ecclesiastical, or social. He was an early utilitarian and a critic of retributive justice with its emphasis on punishing past behavior rather than looking to the future consequences of incarceration. Godwin was also an Enlightenment thinker who advocated disinterested benevolence and had "aspirations for a deliberative culture focused around the exercise of private judgement and the pursuit of truth and justice" Philp (2020, p. 16).

His outlook was utopian, believing that government and artificial constraints imposed by society would eventually become unnecessary as a morality based on reason becomes predominant. This tenet would later play a fundamental role in British Romantic fiction and in Godwin's own flaunting of social conventions. Godwin wrote a number of novels that are read and studied to this day, the most famous being *Caleb Williams*. This novel embodies many of the precepts advanced in *An Enquiry Concerning Political Justice*.

Edith Wharton (1862–1937) (née Edith Newbold Jones) was born into a prominent upper class New York family during the American Civil War.[11] Wharton is best known for her novels and short stories, but she also published poetry and wrote extensively on travel, architecture, and design. Wharton has the distinction of being the first woman to win the Pulitzer Prize for Literature, a full membership in the American Academy of Arts and Science, and an Honorary Doctor of Letters degree from Yale University. She was made a Chevalier of the French Legion of Honor in 1916 for her humanitarian contributions during World War I. Her 1885 marriage to Teddy Wharton was unfulfilling and grew strained over time, ending in divorce in 1913. Wharton spent much of her childhood in France, Italy, Germany, and Spain. After her marriage, she divided her time between Europe and the United States, moving permanently to France after her divorce.

Wharton's literary interests and talents were evident from an early age but were actively discouraged by her mother. She was raised in conformity with the prevailing view that the role of a women from her class was to be a wife and mother, not to pursue a career. Like Godwin, Wharton rebelled against the mores that governed her society. Many of her major novels, such as *The House of Mirth* (1905), *The Custom of the Country* (1913), and *The Age of Innocence* (1920), chronicle the limitations placed on ambitious women whose families belonged to Gilded Age New York society and the adverse effects that the customs of this society had on them.

3. Sequential Decision Problems, Forward-Looking Planning, and Backwards Induction

An *individual sequential decision problem* is modeled using a decision tree. Attention is restricted to decision problems that terminate after a finite number of decisions and concern a decision-maker who is cognizant of all past decisions whenever a choice must be made.[12] For such problems, a *decision tree* consists of a set of nodes, branches, payoffs, and probabilities for the

resolution of any exogenous event.[13]

There are three kinds of nodes. A *decision node* designates a stage in the problem at which the decision-maker makes a choice. A *terminal node* indicates that there are no further decisions to be made. A *chance node* indicates that some exogenous uncertain event takes place.

The problem begins with a decision node called the *initial node*. Associated with each terminal node is the *payoff* to the decision-maker if this node is reached. These payoffs are typically expressed in terms of von Neumann and Morgenstern (1944) utilities although sometimes monetary payoffs are used instead.

Each decision node has one or more *branches* that connect this node to exactly one other node, its *successor*. These branches represent the possible choices available to the decision-maker at this stage of the problem. A chance node is modeled as a decision node for a fictitious player—nature. The branches at a chance node correspond to the possible resolutions of the uncertainty, each with a fixed probability of being chosen. A decision tree contains no cycles. Consequently, for each node distinct from the initial node there is a unique path from the initial node to it. Which decision node the decision-maker is at can be identified by this path.

When, as has been assumed, there is perfect recall, each decision and chance node marks the beginning of a subproblem that can be analyzed in isolation. Such a node together with its successor nodes, branches, and payoffs constitutes a *subtree* of the full decision tree.

These definitions are illustrated in Figure 1. There are four decision nodes, one chance node, and six terminal nodes. The initial node is indicated by a circle and the other nodes by a dot. The chance node is labelled with an N (nature). The terminal nodes are labelled with the payoff obtained if this node is reached. Each branch following a decision node is labelled with a letter and each branch following the chance node is labelled with the probability that that branch is chosen. Each decision and chance node is the initial node of a subtree. At the initial node of the full decision tree, there is a choice between L and R. If L is chosen, the decision-maker has a choice between a and b. However, if R is chosen instead, natures moves picking the left branch with probability .6 and the right branch with probability .4. Following an initial choice of L, if b is chosen, there are no further decisions, but if a is chosen, there is a final choice between c and d. If nature chooses the left branch, the decision-maker has a final choice between e, f, and g, whereas if nature chooses the right branch, the decision-maker has a final

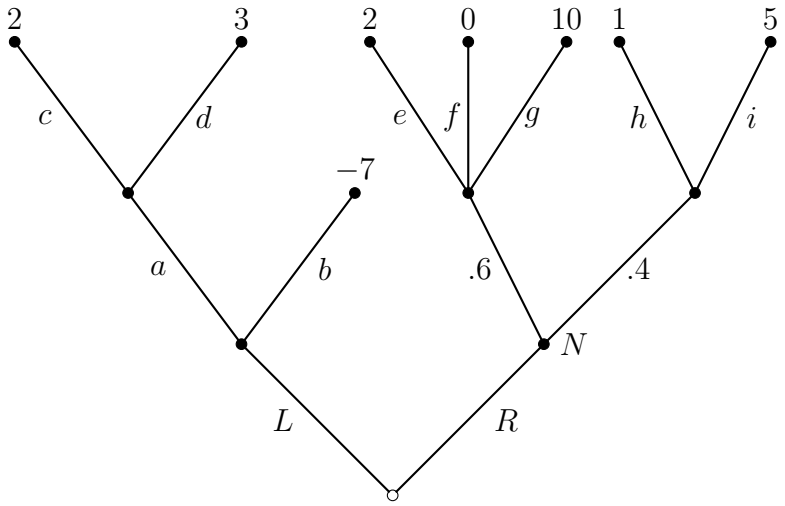
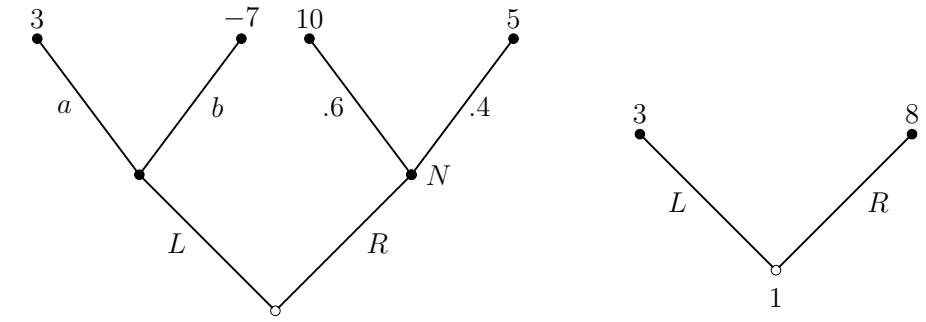


Figure 1: A decision tree.

choice between h and i. Once a final decision is made, the payoffs are realized. For example, a payoff of 3 is obtained by following the path characterized by the sequence of decisions L, a, and d.

A *strategy* specifies what the decision-maker will choose at each of his decision nodes. Because the decision-maker knows the whole structure of the decision tree, he can choose a strategy before actually implementing any of his choices. In effect, a strategy serves as a set of instructions specified in advance saying what to do at each decision node should it be reached. By choosing a strategy in this way, the decision-maker engages in forward-looking contingent decision-making.

In order to determine which strategy maximizes the decision-maker's payoff, forward-looking planning is essential as the best choice at each decision node depends on what the decision-maker anticipates that he and nature will choose at each of the subsequent decision and chance nodes. This is a complex problem, one that *backwards induction* provides a way of solving. Backwards induction is based on the observation that no forward planning is needed at any decision node that immediately precedes a terminal node because, at such a node, the decision-maker only needs to determine which branch (or branches) leads to the highest payoff. Once the choices at these nodes have been determined, each of the subtrees starting with one of these nodes can



(a) Step 1. (b) Step 2.

Figure 2: The backwards induction solution in Figure 1.

be replaced by a terminal node with the payoff received if the choice is made optimally. Sequentially repeating this procedure, eventually the problem is reduced to one in which there is only a single decision node—the initial node. A strategy determined in this way is the *backwards induction solution* to the decision problem being considered. No other strategy can result in a larger expected payoff when viewed from the perspective of the beginning of the decision tree. Thus, backwards induction provides a way of optimally deciding in advance what to do in every contingency. Informally, the decision-maker is forward looking but reasons backwards.

The backwards induction procedure can be illustrated with the decision tree shown in Figure 1. There are five decision nodes in this tree and a strategy specifies what choices are made at each of them. To determine the backwards induction strategy, the choices made at the three decision nodes that are only followed by terminal nodes are considered first. When the choice is between c and d, the latter results in a higher payoff, and so will be chosen. Similarly, g is chosen when the options are e, f, and g, and i is chosen when the options are h and i. By replacing each of these three decision nodes with a terminal node whose payoff is the one received if the payoff maximizing branch is chosen, the decision tree shown in the left-hand panel of Figure 2 is obtained. This process is now repeated at the two decision and chance nodes that are only followed by terminal nodes. At the decision node, a is chosen. When combined with nature's probabilistic choice, this results in the decision tree shown in the right-hand panel of Figure 2, where the payoff of 8 is the expected value of nature's choice. There is only decision node (the

initial decision) in this decision tree. At it, the decision-maker chooses R.

Thus, the backwards induction solution to this problem is the strategy consisting of the contingent choices R, a, d, g, and i. This strategy has an *ex ante* expected payoff of 8. By following this strategy, the contingent choices of a and d are never implemented and which of g and i is implemented depends on nature's choice. Nevertheless, knowing what would have been chosen had the corresponding decision nodes been reached helps the decision-maker determine that he best furthers his objective by first choosing R and then choosing g or i depending on whether nature chooses the left or right branch.

The backwards induction solution describes the decision-maker's plans before any actual choices have been made. If this plan is implemented, R is first chosen, which results in nature making a move. If nature picks the left branch, g is chosen, whereas if nature picks the right branch, i is chosen. This sequence of choices identifies two paths from the initial node to two of the terminal nodes. Which of them is followed depends on nature's choice.

4. Wharton and Forward-Looking Decision-Making

4.1. Wharton's "The Last Asset"

"The Last Asset" (Wharton, 1904) opens with the narrator, Paul Garnett, an American newspaper reporter based in London, seated in a modest restaurant on the Left Bank of Paris. In response to a note he has received from Mrs. Samuel Newell requesting that he call on her at the Ritz Hotel, Garnett exclaims, "The Devil!" (p. 150) Also present is a regular customer, an older American gentlemen who has seen better days. Garnett has conversed with him on previous occasions at this restaurant without having introduced themselves. The conversation that ensues between the two opens with an exchange in which Garnett remarks that the devil may be the source of the unexpected, to which the other responds that "it's generally a woman who's at the bottom of the unexpected. Not ... that that precludes the devil being there too." (p. 150) He then (p. 151) proceeds to offer the following advice to Garnett:

> "Get your life down to routine—eliminate surprises. Arrange things so that, when you get up in the morning, you'll know exactly what is going to happen to you during the day—and the next day and the next. I don't say it's funny—it ain't. But it's

better than being hit on the head by a brick bat. That's why I always take my meals at this restaurant. I know just how much onion they put in things—if I went to the next place I shouldn't. And I always take the same streets to come here—I've been doing it for ten years now. I know at which crossings to look out—I know what I'm going to see in the shop-windows. It saves a lot of wear and tear to know what's coming. For a good many years I never did know, from one minute to another, and now I like to think that everything's cut-and-dried, and nothing unexpected can jump out at me like a tramp from a ditch."

He paused calmly to knock the ashes from his cigar, and Garnett said with a smile: "Doesn't such a plan of life cut off nearly all the possibilities?"

The old gentleman made a contemptuous motion. "Possibilities of what? Of being multifariously miserable? There are lots of ways of being miserable, but there's only one way of being comfortable, and that is to stop running round after happiness. If you make up your mind not to be happy there's no reason why you shouldn't have a fairly good time."

Years earlier, when the wealthy Mr. Newell had suffered ruinous financial reverses, Mrs. Newell and her daughter, Hermione, left him to live among fashionable society in England, where they lived off of the generosity of their acquaintances. When Garnett calls on Mrs. Newell, he is surprised to find that Hermione is engaged to a young French aristocrat. This announcement is followed with a conversation between Mrs. Newell and Garnett that includes the following exchange (pp. 156, 158):

"I was given to understand at once that my husband must appear—if only to establish the fact that we're not divorced."

"Ah—you're not, then?" escaped from Garnett.

"Mercy, no! Divorce is stupid. They don't like it in Europe. And in this case it would have been the end of Hermy's marriage. They wouldn't think of letting their son marry the child of divorced parents."

"How fortunate, then —"

"Yes; but I always think of such things beforehand. And of course I've told them that my husband will be present.""

Mrs. Newell knows that her husband lives in Paris. She asks Garnett to find him and get him to consent to showing up at the wedding, fearing that if she made the request herself that he would refuse. Garnett reluctantly agrees to do so, surmising that Mrs. Newell has nearly exhausted her opportunities for maintaining her lifestyle and that she and Hermione will have a dim future if the wedding is called off.

After some investigation, Newell obtains an address for Mr. Newell near the restaurant he frequents. He encounters the older gentleman there and asks for directions, at which point the latter reveals that he is Mr. Newell. However, he does not want to help his spouse out and asks for a day to think about the request. After his wife and daughter are informed of the situation, Hermione privately tells Garnett that she knows that her father was badly hurt when her parents separated and that she does not want him to be forced to come to the ceremony. When Garnett repeats this conversation to Mr. Newell, he agrees to participate and the wedding takes place, thereby rehabilitating Mrs. Newell's social status and ensuring Hermione's future happiness. Garnett, however, is disgusted at how he has been used by Mrs. Newell and that she has treated her husband "as the last stake in her game, the last asset on which she could draw to rebuild her fallen fortunes." (p. 168) But he is consoled by the fact that he has reconciled father and daughter and has helped facilitate her marriage.

4.2. Forward-Looking Decision-Making in "The Last Asset"

Although ambiguous in some ways, the advice that Samuel Newell offers to Paul Garnett contains the rudiments of forward-looking contingent planning in a sequential decision problem. Garnett is being told to plan in advance what sequence of decisions to make in the future. In modern terminology, this sequence constitutes a path through the decision tree that the decision-maker is facing. As such, the advice advocates adherence to Savage's principle of "looking before you leap" rather than "crossing each bridge when you get to it." It is the former principle that underlies the concept of a strategy.

However, it would be stretching the interpretation of what Wharton puts into the mouth of Samuel Newell too much to say that this passage advocates planning in advance what to do in every possible contingency (i.e, at every decision node), as required by a full-fledged strategy. When there are no uncertain events, implementing a strategy results in one particular path through the decision tree being followed. A strategy also specifies what to

choose at the decision nodes not on this path. It does not seem that Newell is recommending that Garnett plan in advance what to do in those contingencies as well, so Wharton has not provided an account of a strategy as now understood in decision theory. Rather, she provides an informal statement of one of the distinguishing features of a strategy—plan in advance the sequence of decisions that will actually be made.

Although Wharton does not use the terminology of the "look before you leap" aphorism, she is clearly appealing to it. An aphorism is an informal phrase that is meant to capture a general truth in a catchy way. Because of its informality, the scope of the truth that it is meant to represent is often not clear. This is the case with "look before you leap." Nothing in the statement of this aphorism suggests that it is to apply to a sequence of future decisions rather than to a single future choice. Wharton sharpens its advice by explicitly applying it to a sequence of decisions, as in formal models of sequential decision-making. This is evidenced when Newell says, "Get your life down to routine—eliminate surprises. Arrange things so that, when you get up in the morning, you'll know exactly what is going to happen to you during the day—and the next day and the next." It is primarily for this reason that Wharton should be credited as being an early contributor to decision theory.

Garnett's response to Newell that "Doesn't such a plan of life cut off nearly all the possibilities?" can be plausibly interpreted in formal terms in two ways. First, planning a path through a decision tree in advance does preclude making choices at decision nodes off this path. Second, Garnett could simply be suggesting that by waiting until a decision must be made, one leaves open the possibility of making different choices on whatever path is followed. In other words, "crossing each bridge when you get to it" keeps one's options open. Newell's rejoinder is consistent with either possibility but the latter is more likely as it does not involve the more sophisticated idea of making choices in situations that will never be realized.

It is difficult to say with any confidence exactly what Wharton means by her remarks about surprises and uncertainty. On one reading, she could be using Newell to say that one should plan one's life so as to eliminate all uncertainties. However, it is impossible to live such a life. On another reading, she could be advocating planning what to do if a contingency outside one's control eventuates. In her example, plan what to do should a tramp jump out of a ditch. It is also possible to interpret Newell's remark about not knowing how much onion he would get in his meal at a different restaurant

as saying that that is a source of uncertainty that he prefers to avoid. A modern decision theorist would distinguish between these possibilities.

The conversation about planning that opens Wharton's story anticipates the fundamental role that planning plays in her narrative. Mrs. Newell knows that when it comes time for her daughter to marry, in order for the match to be a good one that furthers her social ambitions, she needs never to have been divorced. At the time of her separation, had Mrs. Newell not been so forward-looking, she might have sought a divorce. That might have been what she chose to do if she "crossed each bridge when she got to it." However, she is instead a sophisticated forward-looking decision-maker, so instead "looks before she leaps," as Savage advocates for small-world problems.

5. Godwin and Backwards Induction

5.1. Godwin's 1832 Preface

Caleb Williams is one of the classics of English literature. Five editions were published in Godwin's lifetime and it has been in print ever since. As noted earlier, in the Preface to the 1832 edition of his novel, *Fleetwood*, Godwin set out how he composed *Caleb Williams*. In order to substantiate my claim that Godwin employed a rudimentary from of backwards induction to plan the sequence of events in this novel, it is best to begin by quoting (in part) what Godwin said about his compositional method.[14]

> I formed a conception of a book of fictitious adventure, that should in some way be distinguished by a very powerful interest. Pursuing this idea, I invented first the third volume of my tale, then the second, and last of all the first. I bent myself to the conception of a series of adventures of flight and pursuit; the fugitive in perpetual apprehension of being overwhelmed with the worst calamities, and the pursuer, by his ingenuity and resources, keeping his victim in a state of the most fearful alarm. This was the project of my third volume.
>
> I was next called upon to conceive a dramatic and impressive situation adequate to account for the impulse that the pursuer should feel, incessantly to alarm and harass his victim, with an inextinguishable resolution never to allow him the least interval of peace and security. This I apprehended could best be effected by

a secret murder, to the investigation of which the innocent victim should be impelled by an unconquerable spirit of curiosity. The murderer would thus have a sufficient motive to persecute the unhappy discoverer, that he might deprive him of peace, character and credit, and have him for ever in his power. This constituted the outline of my second volume.

The subject of the first volume was still to be invented. To account for the fearful events of the third, it was necessary that the pursuer should be invested with every advantage of fortune, with a resolution that nothing could defeat or baffle, and with extraordinary resources of intellect. Nor could my purpose of giving an overpowering interest to my tale be answered, without his appearing to have been originally endowed with a mighty store of amiable dispositions and virtues, so that his being driven to the first act of murder should be judged worthy of the deepest regret, and should be seen in some measure to have arisen out of his virtues themselves. It was necessary to make him, so to speak, the tenant of an atmosphere of romance, so that every reader should feel prompted almost to worship him for his high qualities. Here were ample materials for a first volume.

I felt that I had a great advantage in thus carrying back my invention from the ultimate conclusion to the first commencement of the train of adventures upon which I purposed to employ my pen. An entire unity of plot would be the infallible result; and the unity of spirit and interest in a tale truly considered, gives it a powerful hold on the reader, which can scarcely be generated with equal success in any other way.

I devoted about two or three weeks to the imagining and putting down hints for my story, before I engaged seriously and methodically in its composition. In these hints I began with my third volume, then proceeded to my second, and last of all grappled with the first. I filled two or three sheets of demy writing-paper, folded in octavo, with these memorandums. They were put down with great brevity, yet explicitly enough to secure a perfect recollection of their meaning, within the time necessary for drawing out the story at full, in short paragraphs of two, three, four, five, or six lines each.

I then sat down to write my story from the beginning.

In commenting on the 1832 Preface, Dumas (1966, p. 596) says that

> Godwin, thinking perhaps that his readers would be familiar with the 1794 Preface (which had formed part of nearly every edition of *CW* since the second), ignores completely his original political impulse. It is remarkable, too, that although he writes at some length on literary considerations such as unity, plot, and suspense, he neglects to consider specifically how he worked out the novel's catastrophe.

This neglect complicates the assessment of the extent to which Godwin actually employed the backwards construction described in his 1832 Preface.

The novelty of Godwin's method of composition was remarked upon by Charles Dickens in a letter he sent to Edgar Allan Poe in 1845 (Dickens, 1974, pp. 106–108l). Dickens wrote:

> Apropos the "construction" of Caleb Williams. Do you know that Godwin wrote it *backwards*—the last Volume first—and that when he had produced the hunting-down of Caleb, and the Catastrophe, he waited for months, casting about for a means of accounting for what he had done. (emphasis in the original, editorial footnotes omitted)

Poe (1845a, p. 167) paraphrased this passage from Dickens' letter, going on to say that"[t]his mode cannot surely be recommended, but evinces the idiosyncrasy of Godwin's mind." He later used the same passage in his influential essay, "The Philosophy of Composition" (Poe, 1846, p. 164), in order to set the stage for an elucidation of the method he used to compose his poem, "The Raven" (Poe, 1845c).[15]

It is thus safe to infer that the influence of Godwin's form of backwards induction was making itself known in literary circles slightly over a decade after he wrote his 1832 Preface. It would have to wait for over forty years for this kind of reasoning to be formalized by Cayley.

5.2. The Plot of Caleb Williams

Godwin's 1832 Preface provides a brief sketch of the plot of *Caleb Williams*. In order to evaluate the extent to which Godwin used backwards induction in the composition of this novel, it is useful to see how he fleshed out this sketch by providing an extended summary of the major events in the narrative.

The two main protagonists are Ferdinando Falkland, a member of the British landed gentry, and his secretary, Caleb Williams, who Falkland hires as a young man. The events in the novel are narrated by its eponymous title character. Falkland has a chivalrous conception of honor, valuing his good name above all else. He is highly regarded for his benevolence and social position. Like Godwin, Williams places a high value on his independence and his right to live a life free of oppression. He also exhibits a strong degree of curiosity. Both men go to great lengths to strive for what they hold dear—honor and reputation in the case of Falkland, justice and equity in the case of Williams. Each of the novel's three volumes builds up to a climax, what Godwin refers to as a "catastrophe."

Volume I begins with an account of the hiring of Williams as the secretary of the reclusive Falkland and, as a consequence of his curiosity, Williams' accidental discovery of Falkland in a state of anguish examining the contents of a trunk. Falkland's passions get the best of him, accusing Williams of spying on private matters so as to ruin him and threatening Williams with retribution only to appear to later relent. Williams soon realizes that secrecy is expected of him, but is curious as to the nature of Falkland's secret.

The rest of Volume I is devoted to a recounting by Williams of the events that led up to Falkland's reclusivity as related to him by Falkland's steward, Mr. Collins. According to Collins, the esteem with which Falkland was held on his return home from an extended absence on the Continent incited the jealousy of his neighbor, Squire Tyrrel. When Tyrrel observed the partiality to Falkland of his cousin and ward, Miss Melville, he sought vengeance by confining her to his house and commanding her to marry a boorish youth. Miss Melville's refusal was followed by her flight from Tyrrel's home and rescue by Falkland. During her subsequent fatal illness, Tyrrel ordered her arrest for supposedly owing him for her past maintenance. The assistance that Falkland provided Miss Melville during her flight and illness further raised Tyrrel's antipathy towards him. Tyrrel was also indignant that his tenant, Mr. Hawkins, refused his offer to employ Hawkins' son. In retaliation, Tyrrel did all he could to deprive Hawkins of his livelihood. After Hawkins' son was observed on Tyrrel's property removing obstacles that had been placed to hinder Hawkins' access to his land, Tyrrel had the son charged of a crime that carries the death penalty, which led the two Hawkinses to go into hiding. Because of his ill treatment of Miss Melville, Tyrrel was condemned by the community and banned from the local nobility's assembly. Nevertheless, he went there and physically assaulted Falkland, thereby publicly insulting

what Falkland holds most dear, his honor. Soon thereafter, Tyrrel was found nearby stabbed to death.

In the climax to Volume I, Falkland is charged for the murder but is acquitted because it is felt that such a heinous act is out of character for someone who is so honorable. The volume ends with blame shifted to the Hawkinses, who are then convicted and executed. From this point forward, Falkland withdraws from public life to become the recluse that Williams encountered when he was hired.

Volume II begins with a detective story that chronicles Williams' investigation into the nature of Falkland's secret.[16] He begins his enquiries by asking Falkland seemingly benign questions and entering into discussions that are designed to lure Falkland into revealing his secret. Soon after Williams discovers a letter from Hawkins that leads him to question Hawkins' guilt for the murder, Falkland confronts Williams about prying into his personal affairs. Falkland confides that he is miserable not only for having his honor besmirched, but also for not having any opportunity for avenging it. After Williams expresses remorse and his veneration for Falkland, he asks to leave Falkland's service, which only enrages his master.

When a peasant is examined on a charge of murder by Falkland in his capacity as a justice of the peace, Falkland is so overcome by emotion that he flees the proceedings leading Williams to surmise that Falkland is Hawkins' murderer and that evidence to that effect lies in Falkland's trunk. When Falkland catches Williams breaking into the trunk, Falkland confesses to the murder, saying that he sacrificed his peace of mind and the Hawkinses so as to maintain his public regard. He swears Williams to secrecy.

During an extended visit by Falkland's brother, Mr. Forester, Falkland becomes perturbed by the intimacy that arises between them. He accuses Williams of trifling with him and threatens Williams with an inescapable vengeance if he should cross him. Williams later visits Forester at his home and confides that in spite of his high regard for Falkland, he is miserable in his service. Williams refuses to reveal the reasons for this misery (the secret) but nevertheless Forester offers to shelter him if he ever needs assistance. They are interrupted by Falkland, who suspects Williams of duplicity.

Williams then submits to Falkland a request to leave his service. Falkland rejects the request and vows to destroy Williams if he attempts to quit him. Williams then decides to escape, first hiding his possessions that were not portable. He is overtaken and given a letter from Forester encouraging him to return to clear his name, offering support if he is successful in doing so. In the

meantime, Falkland has discovered Williams' possessions. He plants money and other valuables with them, convincing Forester that Williams stole them from him. Williams is imprisoned but faces a lengthy wait for his trial. Following an unsuccessful attempt to escape, Williams is visited by Falkland's footman, Thomas, who is is appalled by the conditions of William's cell. Volume II concludes with Thomas later surreptitiously providing tools for a jail break and Williams' subsequent successful escape.

The first part of Volume III is devoted to the pursuit and capture of Williams. Following his escape, Williams is robbed and beaten by three members of group of thieves, one of whom, Grimes (called Jones in the first edition) is talked out of killing him by his compatriots. Williams is found by the leader of the gang and provided shelter during his recovery. At the leader's urging, Grimes is expelled because of his brutality to Williams. As vengeance, Grimes turns into Williams' pursuer in the hopes of obtaining the substantial reward offered for his capture. In spite of frequent changes of locale and disguise, by circulating information that suggests that Williams is a dangerous criminal, Grimes is successful in tracking Williams' whereabouts. When Williams is brought before a magistrate, he protests his innocence of the robbery and reveals Falkland's secret to no avail. However, at Williams' trial for the alleged robbery, nobody appears as a witness against him, so he is set free.

After the trial, Falkland informs Williams that he had employed Grimes to keep tabs on him and to prevent his capture. He asks Williams to help restore his reputation by signing a paper attesting that he is innocent of Tyrell's murder, which Williams refuses to do. Falkland vows to use his formidable power to crush Williams, which he does by having Grimes ensure that Williams' reputation is sullied wherever he settles. As Williams is about to depart England in the hope of escaping from Falkland's reach, Grimes confronts Williams and informs him that he is free to live within the British Isles but will be arrested if he makes another attempt to flee overseas.

Williams then goes to a magistrate near Falkland's home, relates Falkland's secret, and requests that he be charged with murder. At the commencement of Falkland's trial, when Williams sees how frail Falkland is, he regrets having pressed his charge but feels compelled to tell his story so that the truth will be revealed. On finishing his story (which is interspersed with eulogies to Falkland's virtues), Williams goes on to say that he now believes that his own behavior towards Falkland is responsible for Falkland's decline, that he is in effect Falkland's murderer. Falkland then confesses to the mur-

der of Tyrrel, saying that he now recognizes that Williams is the virtuous one and that as a result of his momentary lapse when he committed the murder, he has subsequently lead a life of cruelty so as to preserve his reputation. Falkland dies three days later leaving Williams to declare that he began his memoirs in order to vindicate his character but that now he is miserable and has no character to preserve. Nevertheless, he believes that Falkland's life provides a valuable lesson about the corrupting influence of false beliefs and societal baseness on an inherently virtuous person, a lesson that justifies completing his account.

5.3. Philosophy and Psychology in Caleb Williams

Graham (1990) notes the importance of the two parts of the title of Godwin's novel: "Things as They Are" and "The Adventures of Caleb Williams." Godwin's 1832 Preface stresses the adventure story and only alludes to the main title by some brief remarks about the mental states and motivations of its protagonists and of the tyranny to which the victim is subjected. In the 1794 Preface written for the first edition of *Caleb Williams* but not published until it was included in the 1796 second edition, Godwin (2005, p. 3, emphasis in the original) sets out his goal of elucidating "Things as They Are" by depicting the pernicious effects of government in contributing to the despotism that the citizenry face.[17]

> The following narrative is intended to answer a purpose more general and important than immediately appears upon the face of it. The question now afloat in the world respecting *Things As They Are* is the most interesting that can be presented to the human mind. While one part pleads for reformation and change, the other extols in the warmest terms possible the existing constitution of society. It seemed as something would be gained for the decision of this question, if that constitution were developed in its practical effects. ... It is now known to philosophers that the spirit and character of the government intrudes itself into every rank of society. But this is a truth worthy to be communicated to whom books of philosophy and science are never likely to reach. Accordingly, it was proposed, in the invention of the following work, to comprehend, as far as the progressive nature of a single story would allow, a general review of the modes of domestic

> and unrecorded despotism by which man becomes the destroyer of man.

Here, Godwin contrasts the radical elements in British society who advocated for revolutionary change and the forces of conservatism epitomized by Edmund Burke in his *Reflection on the Revolution in France* (Burke, 1790). While critical of the apologists of the existing order, "Godwin's central distaste for violent revolutionary activity [contrasts with his advocacy of] the tranquil operation of reason leading to gradual improvement" in social institutions and relations (Clemit, 1993, p. 41).[18]

When Godwin disparagingly says that "government intrudes itself into every rank of society," he is alluding to his view that the characters (i.e., the dispositions) of all ranks in society are shaped by the form of government that they live under. Graham (1990, p. 152) encapsulates this view when he says that Godwin's "*Political Justice* emphasizes the role of government in encouraging and perpetuating the ideologies that contribute to its power and authority." But as Graham (1990, p. 28) also notes, Godwin also believes that governmental institutions (such as the judicial and penal systems) have a more direct pernicious effect, as illustrated by the psychological effects on Williams of his confinement in prison

Godwin does not provide a clear statement of the purpose he alludes to in the first sentence of his 1794 Preface. Myers (1972, p. 597) contends that "[s]ince for Godwin political problems are essentially moral problems, ... the valuable lesson he wishes to teach is primarily a moral one." In a response to one of his critics, Godwin (1795, p. 94) provides some clarification that his purpose is "to expose the evils which arise out of the present system of civilized society; and, having exposed them, to lead the enquiring reader to examine whether they are, or are not, as has commonly been supposed, irremediable ..." He goes on to say that his concern is with "the administration of justice and equity, with its consequences, as it exists in the world at large, and in Great Britain in particular" rather than with the content of English law *per se*.

In his 1832 Preface, Godwin (2005, p. 351) says that he started writing *Caleb Williams* in the third person, but then switched to a first person narrative. He went on to say that this perspective

> was infinitely the best adapted, at least, to my vein of delineation, where the thing in which my imagination revelled the most freely,

> was the analysis of the private and internal operations of the mind, employing my metaphysical dissecting knife in tracing and laying bare the involutions of motive, and recording the gradually accumulating impulses, which led the personages I had to describe primarily to adopt the particular way of proceeding in which they afterwards embarked.

While Godwin hoped that reason would prevail in the quest for a more just society, increasingly over time, he recognized its limits. This is reflected by the central role that the false beliefs Falkland and Williams hold about each other feature in the narrative of *Caleb Williams*. Philp (2020, p. 77) provides a succinct statement of the possible tension between beliefs and reason:

> In *Caleb Williams*, Godwin first fully explored the extent to which opinion could be something that resisted being brought under the command of reason. Prejudice and presupposition repeatedly undo Williams and he and Falkland are unable to reason together because of the societal prejudices (and inequalities) that they have in various ways inherited. For Godwin this was a growing recognition of the difficulties of winning the argument by appealing to reason alone. His gradualism, his suspicion of political activity and associations and his own contained and careful deliberative manner all pointed to his acknowledgement of the challenges that people face in commanding their faculties so that reason and rationality would direct their beliefs, rather than opinion, rumour, interest, emotion and prejudice.[19]

We thus see that Godwin not only intended to write a gripping adventure story, he also wanted to communicate what he regarded as important insights about moral and political philosophy and psychology to those who do not read philosophical treatises. Myers (1972, p. 591) concludes that

> the political reflections and psychological analyses cannot be considered in isolation from the moral pattern that forms the core of the romance. Godwin's statements about the book emphasize different aspects at different times: neither of the two major Prefaces is completely accurate as a guide to the work.

5.4. Is Godwin's Stated Compositional Method Backwards Induction?

To determine whether Godwin employed backwards induction in the composition of *Caleb Williams*, two questions need to be distinguished. First, is the method of composition described in his 1832 Preface really a form of backwards induction? Second, did Godwin actually use this method of composition? These two questions are taken up in turn.

When backwards induction is used to solve a sequential decision problem, the choice to be made at every decision node should it be reached is determined. In the decision tree depicted in Figure 1, this requires making a choice at each of the five decision nodes. While employing backwards induction ensures that the maximal *ex ante* expected payoff is achieved, it is not necessary to determine a complete strategy in order to achieve this payoff. Instead, the decision-maker could first identify what the maximal expected payoff is and then determine what sequence of decisions results in this outcome. If there are no chance moves, this simply involves choosing one of the terminal nodes with the highest payoff and then determining what sequence of decisions terminates at this node. If there are chance nodes, the analysis is somewhat more complicated but, in general, it also does not require specifying a complete strategy. For example, in Figure 1, the maximal expected payoff of 8 is obtained if g and i are chosen following the choice of the left and right branch, respectively, by nature. To achieve this expected payoff, R must be chosen at the initial node. It is not necessary to determine what to choose in the subtree following a choice of L because whatever is chosen is suboptimal. That is, it is dominated by some other achievable expected payoff. Put another way, backwards induction is used to determine an optimal strategy in the subtree beginning with nature's move. Because the maximal expected payoff of 8 in this subtree is larger than any of the payoffs following a choice of L, the decision-maker knows that he needs to first choose R. This simplified procedure uses backwards reasoning but it does not, and need not, specify a choice at every decision node. While this procedure is not strictly speaking backwards induction, it retains the basic features of backwards induction, so it is reasonable to say that it exemplifies backwards induction reasoning.

The method of composition described in Godwin's 1832 Preface concerns a sequential decision problem that does not involve any chance moves. Consequently, Godwin only needed to first determine what outcome best serves his purpose and then to figure out how to get to it. As has been discussed

above, it is not completely clear what Godwin's ultimate objective was. But Godwin is very explicit in this Preface that he plotted his novel backwards "from the ultimate conclusion to the first commencement of the train of adventures." Moreover, this method of composition "gives it a powerful hold on the reader, which can scarcely be generated with equal success in any other way." On one level, Godwin may only be claiming that the use of this method is best suited for crafting a series of adventures that will capture the interest of his readers. On another level, Godwin may be claiming that it best serves his purpose in exposing "Things as They Are." It thus seems reasonable to conclude that Godwin's method is a version of backwards induction. While Godwin does not specify what to do in every possible contingency, he does reason backwards in order to determine how to achieve his desired outcome, and that is the essence of backwards induction.

A noteworthy feature of Godwin's description is that he uses backwards reasoning twice. He first used backwards reasoning to sketch the basic plot. Then, he applied this method once more in order to work out more of the details of the plot before he began composing the novel in the conventional order. Thus, his method is applied at different levels of specificity. When considering the basic structure of the plot, backwards reasoning is employed. At the next level, when a more detailed outline of the events in the novel is set out, the three volumes are again considered in reverse order. Godwin is silent as to whether as he prepared an outline for each volume at this stage, he employed backwards reasoning, plotted the sequence of events in a volume from beginning to end, or used some combination of both methods. However, Godwin claimed that once this preparatory stages were completed, he wrote his narrative "from the beginning."

5.5. Did Godwin Actually Use Backwards Induction?

Authors are not alway candid when discussing their own works works, so even if it is agreed that the method of composition described in Godwin's 1832 Preface is a form of backwards induction there remains the question of whether the composition of *Caleb Williams* proceeded in the way that Godwin alleges. Fortunately, a great deal of evidence exists that sheds insight into this question. The holograph manuscript of this novel used by the printers is in the Forster Collection held by the Victoria and Albert Museum (Godwin, 1793–94).[20] In addition, Godwin's diaries are among the Abinger Papers held by the Bodleian Library (Godwin, 1788–1836). In his diaries,

Godwin wrote a few lines each day describing his activities in very abbreviated form. From these entries it is possible to learn whether Godwin wrote or revised text that day, what he read, and who he met with. Digital versions of both the manuscript and the diaries are available for viewing online.

Whether the 1832 Preface accurately describes Godwin's method is complicated by the discovery by Gilbert Dumas (Dumas, 1966) over a century and half after *Caleb Williams* was published that Godwin substituted a radically new ending for the novel as it was in press. Dumas discovered the original ending had been bound with the holograph manuscript. From a comparison of Godwin's diaries and the holograph manuscript, Dumas (1966, p. 580) concludes that the original ending was completed on April 30, 1794 and the new ending was written a few days later between May 4th and 8th. Godwin does not say what he revised that week, but the evidence suggests to Dumas that it was the ending of *Caleb Williams*.[21] The novel was published less than three weeks later on May 26th.

The original ending begins after Williams sees how frail Falkland is at the final trial. In this version, Falkland defends himself by appealing to his reputation and by questioning Williams' motives. Williams responds by saying that his objective is to secure justice. Eventually, the magistrate cuts Williams off and admonishes him for his dishonesty. At this point, two pages of the manuscript are missing. When the narrative picks up, Williams is confined with Gines (called Jones in this version) as his cruel keeper, suffering in the throes of madness. In his lucid moments, Williams completes his memoirs, commenting that "it too plainly appears in my history that persecution and tyranny can never die!" (Godwin, 2005, p. 344) While hopeful that his misfortunes will provide a lesson to others, Williams is fearful that his readers will not be convinced of their soundness but, instead, regard them as perpetuating his shame, which he cannot endure. Apparently drugged, Williams ends his narrative feeling completely defeated, the shell of what was once a man.

One might expect that the replacement of the original ending by the new one would result in a disjointed narrative. This is not the case. As Myers (1972, p. 620) remarks, "[t]he final chapter and the first eight paragraphs of the Postscript which Godwin retains were initially intended as a prelude to one kind of catastrophe in the first ending; yet it is remarkable how smoothly he converts the original material in a preamble for a quite different catastrophe."[22]

A number of conjectures have been made as to why Godwin revised the

novel's ending.[23] Dumas (1966), among others, has argued that the political purpose announced in the 1794 Preface (composing an adventure that illustrates the pernicious effects of tyranny) is compromised by the published ending with its foregrounding of psychological concerns. While acknowledging that the revised ending puts more emphasis on psychological factors than in the original, Myers (1972) and Graham (1990) both regard the new ending as providing a more satisfactory conclusion to the narrative and a better instantiation of the principles put forward in Godwin's *Enquiry Concerning Political Justice*.

For Myers, it is the exploration of these principles that is the moral purpose of Godwin's novel. While the condemnation of tyranny in its many forms is one of these principles, there are others that Godwin wanted to elucidate, such as the principle of impartiality and "the dangerous consequences of egoism, subjectivity, and prejudice" (p. 610). Myers disputes the claim that by rewriting the ending of *Caleb Williams*, Godwin changed the purpose he had for writing his novel. She argues that the fact that Godwin

> substitutes a totally different conclusion, one which is almost diametrically opposite in its view of Caleb and Falkland, indicates his realization that the ending he had set down in his original hints and memorandums for the story was not in accord with the tale as it had developed in the actual writing. (p. 623)

For example, she argues that "the new denouement both completes the basic moral pattern set up in the book and underscores the principle of impartiality which is the root of the moral system elaborated in *PJ*" (p. 601). Furthermore, "[t]he moral principle of impartiality, which the rewritten ending clearly establishes as being of central thematic importance in *CW*, is one to which Godwin's adherence never wavers" (p. 626).

As we have seen, a central tenant of Godwin's philosophy is that deliberation guided by reason is conducive to identifying truth. However, Godwin recognized that this ideal is only partially realized when beliefs are distorted by prejudice and presupposition. Graham (1990, p. 41) argues that the trial and its outcome in the revised ending serve to illustrate this ideal and its limitations.

> The tribunal of truth remains an ideal and the novel demonstrates how the ideal may be corrupted and how it may be approached. The sources of corruption are not limited to prejudice and class

> difference. Corruption enters whenever reason is tainted by passion, objectivity by subjectivity. Falkland and Caleb are changed towards each other by openness and sincerity. For a brief moment at a real trial, truth is spoken because a tribunal at least provides a forum that permits truth to be heard. ... A forum, however corrupt, offers at least an occasion when concealment is resisted, when speaking may be frank, and truth may emerge.

In order to assess the degree to which Godwin employed backwards induction reasoning in the composition of *Caleb Williams*, it is important to distinguish between the planning of the basic plot elements and the actual writing of the manuscript. In his 1832 Preface, Godwin says that the wrote his novel from beginning to end but planned its main narrative structure backwards. A feature of backwards induction is that whatever outcome one decides on, the path to it is determined by backwards reasoning. Once the path is chosen, the resulting plan is implemented in the conventional order. If Godwin had followed this method exclusively, all of the plot details would have been worked out from the ending back to the beginning before he began writing. This is clearly not the case, as his diaries confirm. It would, however, be unrealistic to expect to see backwards induction employed with that much detail in practice.

From the archival sources, there is no reason to doubt that Godwin prepared his outlines as described in his 1832 Preface. However, backwards induction is not merely a method for identifying a path through a decision tree; the path so identified is meant to be implemented. So the question then becomes: To what extent did the completed novel conform to his outlines? The answer to this question is complicated by the fact previously noted that Godwin was not specific about what the novel's catastrophe is. It is also complicated by the fact that Godwin substituted a new ending before publication and made revisions in subsequent editions.[24]

In his 1832 Preface, Godwin seems to suggest that the secret murder would occur at the beginning of Volume II, not as the climax of Volume I. However, even if, as Godwin alleges, the idea of the murder was thought of before the theme for Volume I was devised, for dramatic purposes it makes sense for it to take place at the end of Volume I. It is likely that this is what Godwin actually had in mind when he first conceived of the subject matter for Volume II, so the placement of the murder does not deviate in any significant way from Godwin's outlines.

In all other respects, as my summary of the main plot elements of *Caleb Williams* demonstrates, the published novel accords with the description Godwin retrospectively provided in 1832 regardless of which ending is considered. In particular, the sequence consisting of the murder, its discovery following the investigation by someone curious about its perpetrator, and the subsequent tale of "flight and pursuit" agree with Godwin's account. Furthermore, the traits of the pursuer and victim are realized in Falkland and Williams. The former is wealthy and endowed with "amiable dispositions and virtues," whereas the latter is curious and exhibits "perpetual apprehension" during his pursuit. Of course, in a brief summary, Godwin could not relate all of the main plot elements. Without knowing exactly what Godwin had in mind for the final catastrophe, there is no evidence that Godwin deviated from his original overall plan in any significant way. The exact form of the ending could have been something that Godwin left to be worked out while he wrote, as he did with other plot details.

Godwin's views evolved over time, and this is reflected in the changes he made to his *Enquiry Concerning Political Justice* and *Caleb Williams* in their revised editions.[25] But in evaluating the extent to which Godwin used backwards induction when writing his novel, it is only the composition of the first edition that is relevant. In the few days between writing the withdrawn ending and the one that was published, there is no evidence that Godwin changed his political philosophy in any way. Consequently, it is unlikely that Godwin changed what he was trying to achieve in any fundamental way when he wrote the new conclusion.

If, as Myers (1972) and Graham (1990) suggest, the purpose of *Caleb Williams* is to provide an entertainment that illustrates the principles set forth in Godwin's *Enquiry Concerning Political Justice*, then the concluding catastrophe must be an event that exemplifies them. If Myers and Graham are right, then the revised ending more effectively serves Godwin's didactic objectives than the one that was withdrawn. If this is the case, then the published version of the novel better implements Godwin's original plans than the original draft. If this is so, then the basic plot elements worked out using backwards reasoning were, in fact, successfully realized when the narrative was written from beginning to the revised ending.

Myers (1972, p. 602) remarks that "[a]s with *PJ*, Godwin apparently saw more distinctly into *CW* as the work developed in the actual process of writing." In modern decision-theoretic terminology, he exhibited "growing awareness."[26] Backwards induction would seem not to make any allowance

for a decision-maker to change his mind in this way. However, the phenomenon of growing awareness did not result in Godwin making extensive revisions to the draft he completed in April 1794 before the book was published the next month. The revised ending is only a few pages long and, as remarked by Myers, integrates well with the preceding text.

Even if Godwin's revised ending is taken to be a departure from his original outlines, which we have seen is disputable, it is fair to say that Godwin employed a rudimentary form of backwards induction in planning his novel and made at most one significant adjustment to his plans as he wrote the narrative. Consequently, not only does the method of composition described in the 1832 Preface embody the central features of backwards induction reasoning, for the most part Godwin actually wrote *Caleb Williams* in compliance with the outlines he prepared using this methodology.

6. Conclusion

To the extent that rational choice theory and works of literary fiction offer accurate insights into the reasoning and behavior of individuals, one should expect that literature would be a rich source of illustrations of decision theory's main ideas. Because decision theory is a relatively new discipline, it should also not be surprising to find literary antecedents (if only in very rudimentary form) for decision-theoretic concepts that were only later developed formally. It is one thing to create a narrative that exemplifies principles of decision theory; it is another to articulate what those principles are. It speaks to the genius of Edith Wharton and William Godwin that they were able to do so. In the case of Wharton, she formulated a version of sequential forward-looking decision-making that captures one of the essential features of a strategy in a decision tree. In the case of Godwin, he provided a clear statement of the kind of reasoning that underlies backwards induction. For these reasons, they should be regarded as two of decision theory's pioneers.

Notes

[1]In Weymark (2022), I noted in passing that Wharton's short story provides an informal account of Savage's formal model of forward-looking decision-making but I did not subject her views to any critical assessment. Nor did I comment on a second way that forward planning plays in her story.

[2]Ferguson (1989) discusses some of the early contributions to this problem. Hill (2009) provides a clear exposition of how backwards induction can be used to solve an optimal stopping problem when there is full information about the possible values.

[3]See Rust (2008) for a brief overview of the use of backwards induction in the 1940s and 1950s.

[4]Schwable and Walker (2001) convincingly demonstrate that claims of earlier uses of backwards induction in formal game theory models are incorrect.

[5]There are now dozens of articles and a few monographs devoted to applying game theory to literature, of which the books by Brams (2012) and Chwe (2013) are prominent examples. Chwe (2013, pp. 30–34) provides an overview of some of this literature. An early example in which both of the perspectives on the relationship between game theory and literature described above are present is provided by von Neumann and Morgenstern's analysis of the pursuit of Sherlock Holmes by Professor Moriarty in "The Adventure of the Final Problem" (Conan Doyle, 1893) in terms of a mixed strategy Nash equilibrium (von Neumann and Morgenstern, 1944, pp. 176–178). Read (2020, p. 370) regards Conan Doyle's story and Edgar Allan Poe's "The Purloined Letter" (Poe, 1845b) as being the two "central texts used to discuss how literature and game theory can mutually support one another."

[6]See Hutchinson (1983) and Livingston (1991, pp. 69–75).

[7]See Chwe (2013, Chap. 3) and Wiese (2012).

[8]According to Chwe (2013, p. 1), "Austen's novels do not only provide 'case material' for the game theorist to analyze, but are themselves an ambitious theoretical project, with insights not yet superseded by modern social science."

[9]For references to some of the main works in which Elster has used literature to explore the nature and limits of rational agency, see the bibliography in Livingston (1991). Livingston's *Literature and Rationality* (Livingston, 1991) is his most significant contribution to this subject.

[10]For an overview of Godwin's life and philosophy, see Philp (2021). The critical literature on Godwin's literary output is vast. Clemit (1993) discusses Godwin's main novels in some detail, drawing on her careful study of archival sources. Graham (1990) provides a book-length analysis of *Caleb Williams*.

[11]The biography of Wharton by Hermione Lee (Lee, 2007) considers the influence of Wharton's own experiences on her writings in addition to providing a magisterial account of her life. For a brief introduction to Wharton's life and writing by one of her other biographers, see the 1991 Prologue and Epilogue by Cynthia Griffin Wolff to Wharton's 1900 novel, *The Touchstone* (Wharton, 1991).

[12]The latter property is violated if there is imperfect recall of some past decision or if it is not known what the resolution of some prior uncertain event is.

[13]For introductions to sequential decision problems and their solutions, see Binmore (1992) and Raiffa (1968).

[14]The 1832 Preface has been widely reprinted. The following quotation is taken from the Penguin Classics edition (Godwin, 2005, pp. 349–350).

[15]As noted by the editors of Dickens' letters in footnotes to the passage from the letter quoted above, Poe (1845a) uses Dickens' remarks without attribution and Poe (1846) paraphrases what Dickens wrote even though he purports to directly quote Dickens. Hughes (1977) interprets Dickens as saying that Godwin composed his book backwards rather than that he planned it that way, which is not how Godwin described his own method of composition. E. C. Bentley says that after he had an outline of the plot of his classic detective novel *Trent's Last Case* (Bentley, 1913), he wrote a draft of the last chapter before writing the rest of the novel (Bentley, 1940, pp. 252–254). I owe this observation to Murch (1958, p. 203).

[16] *Caleb Williams* is recognized as being a source for some of the key features of detective fiction. See Murch (1958, pp. 29–35) and Graham (1990, pp. 59–70).

[17]The French Revolution triggered a period of repression in Britain whose most prominent manifestation was a series of trials for sedition or treason of radicals who proposed parliamentary reforms. Just prior to publication, Godwin withdrew his original preface. When it was included in the second edition, Godwin remarked that the withdrawal was made "in compliance with the alarms of the booksellers" and that "[t]error was the order of the day; and it was feared that even the humblest novelist might be shown to be constructively a traitor. (Godwin, 2005, pp. 3–4)

[18]Godwin had mixed feelings about the aristocracy, admiring their talents but blaming corrupt political institutions for diverting these talents from enlightened ends. See, for example, Clemit (1993, p. 40).

[19]Graham (1990, pp. 99–104) explores the social and psychological origins of these challenges and how they feature in *An Enquiry Concerning Political Justice* and *Caleb Williams*.

[20]See Dumas (1966) and Clemit (2017) for valuable commentaries on this manuscript.

[21]Dumas (1966, p. 576) and Myers (1972, p. 598) both note that Godwin never acknowledged in anything that he wrote that the ending was changed.

[22]This is not to say that the earlier parts of the novel could not be revised so as to provide a better prelude for the conclusion. Myers (1972, p 601) suggests that "[t]he additions, deletions, and substitutions made during the revisions of the second and third editions represent Godwin's attempt to bring the rest of the novel into line with the changed vision of the book reflected in the rewritten ending."

[23]See Graham (1990, pp. 36–42) for a critical examination of the main alternatives.

[24]A summary of Godwin's main revisions in subsequent editions of *Caleb Williams* appears in the "Note on the Text" in Maurice Hindle's Penguin Classics edition of the novel (Godwin, 2005).

[25]See Myers (1972) and Philp (2021).

[26]Growing awareness has been modeled in decision theory in various ways. See Karni and Vierø (2017) for a discussion of different ways that unawareness has been formalized.

References

Bellman R (1959) *Dynamic Programming*. Princeton, NJ: Princeton University Press.

Bentley EC (1913) *Trent's Last Case*. London: Nelson.

Bentley EC (1940) *Those Days*. London: Constable.

Binmore K (1992) *Fun and Games: A Text on Game Theory*. Lexington, MA: D. C. Heath.

Brams SJ (2012) *Game Theory and the Humanities: Bridging Two Worlds*. Cambridge, MA: MIT Press.

Burke E (1790) *Reflections on the Revolution in France, And on the Proceedings in Certain Societies in London Relative to that Event. In a Letter Intended to Have Been Sent to a Gentleman in Paris*. London: J. Dodsley.

Cayley A (1875) Problem 4528. In: Miller WJCA (ed.) *Mathematical Questions With Their Solutions, From the "Educational Times," With Many Papers and Solutions Not Published in the "Educational Times."*, volume 23. London: C. F. Hodgson and Son, pp. 18–19.

Chwe MSY (2013) *Jane Austen, Game Theorist*. Princeton, NJ: Princeton University Press.

Clemit P (1993) *The Godwinian Novel: The Rational Fictions of Godwin, Brockden Brown, Mary Shelley*. Oxford: Clarendon Press.

Clemit P (2017) *Caleb Williams*: A description of the holograph manuscript. The Shelley–Godwin Archive. `http://shelleygodwinarchive.org/contents/msl_1876_forster_223/introduction/`.

Conan Doyle A (1893) The adventure of the final problem. *The Strand Magazine* 6: 558–570.

Dickens C (1974) To Edgar Allan Poe, 6 March 1842. In: House M, Story G and Tillotson K (eds.) *The Letters of Charles Dickens, Volume Three: 1842–1843*. London: Clarendon Press, pp. 106–108. Associate Editor: Noel C. Peyrouton.

Dumas GG (1966) Things as they were: The original ending of *Caleb Williams*. *Studies in English Literature, 1500–1900* 6: 575–597.

Elster J (2009) Interpretation and rational choice. *Rationality and Society* 21: 5–33.

Ferguson TS (1989) Who solved the secretary problem? *Statistical Science* 4: 282–289.

Godwin W (1788–1836) *Diaries*. Shelfmarks MS. Abinger e. 1 – e. 33, Bodleian Library, University of Oxford, Oxford. Available online at `https:`

//digital.bodleian.ox.ac.uk/collections/abinger/.

Godwin W (1793) *An Enquiry Concerning Political Justice and Its Influence on General Virtue and Happiness*. London: G. G. J. and J. Robinson.

Godwin W (1793–94) Holograph manuscript of *Caleb Williams*. Shelfmark MSL/1876/Forster/223, Victoria and Albert Museum, London. Available online from the Shelley–Godwin Archive. http://shelleygodwinarchive.org/contents/caleb_williams/.

Godwin W (1794) *Things as They Are; Or, The Adventures of Caleb Williams*. London: B. Crosby.

Godwin W (1795) Correspondence. *The British Critic* 6: 894–95.

Godwin W (1832) *Fleetwood: Or, The New Man of Feeling*. London: R. Bentley. The Standard Novels Edition. Originally published in 1805.

Godwin W (2005) *Things as They Are; Or, The Adventures of Caleb Williams*. London: Penguin Books. Edited with an Introduction and Notes by Maurice Hindle.

Graham KW (1990) *The Politics of Narrative: Ideology and Social Change in William Godwin's* Caleb Williams. New York: AMS Press.

Hill TP (2009) Knowing when to stop: How to gamble if you must—the mathematics of optimal stopping. *American Scientist* 97: 126–133.

Hughes D (1977) The composition of *Caleb Williams*: Dickens's misunderstanding. *Dickens Studies Newsletter* 8: 80.

Hutchinson P (1983) *Games Authors Play*. London: Methuen.

Karni E and Vierø ML (2017) Awareness of unawareness: A theory of decision making in the face of ignorance. *Journal of Economic Theory* 168: 301–328.

Lee H (2007) *Edith Wharton*. London: Chatto & Windus.

Livingston P (1991) *Literature and Rationality: Ideas of Agency in Theory and Fiction*. Cambridge: Cambridge University Press.

Murch AE (1958) *The Development of the Detective Novel*. London: Peter Owen.

Myers M (1972) Godwin's changing conception of *Caleb Williams*. *Studies in English Literature, 1500–1900* 12: 591–628.

Philp M (2020) *Radical Conduct: Politics, Sociability and Equality in London 1789–1815*. Cambridge: Cambridge University Press.

Philp M (2021) William Godwin. In: Zalta EN (ed.) *The Stanford Encyclopedia of Philosophy*, Summer 2021 edition. Metaphysics Research Lab, Stanford University.

Poe EA (1845a) A chapter of suggestions. In: Hale SJ (ed.) *The Opal: A*

Christian Gift for the Holy Days. New York: J. C. Riker, pp. 164–170.
Poe EA (1845b) The purloined letter. In: *The Gift: A Christmas, New Year, and Birthday Present for 1845*. Philadelphia: Carey and Hart, pp. 41–61.
Poe EA (1845c) The raven. *The American Review: A Whig Journal of Politics, Literature, Art and Science* 1: 143–145.
Poe EA (1846) The philosophy of composition. *Graham's American Monthly Magazine of Literature and Art* 28: 163–167.
Raiffa H (1968) *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Reading, MA: Addison-Wesley.
Read D (2020) The five games of Mr Edgar Allan Poe: A study of strategic thought in 'The Purloined Letter'. *Rationality and Society* 32: 369–401.
Rust J (2008) Dynamic programming. In: Durlauf SN and Blume LE (eds.) *The New Palgrave Dictionary of Economics, Second Edition, Volume 2*. Basingstoke and New York: Palgrave Macmillan, pp. 575–593.
Savage LJ (1972) *The Foundations of Statistics*. Second revised edition. New York: Dover. First edition published in 1954.
Schwable U and Walker P (2001) Zermelo and the early history of game theory. *Games and Economic Behavior* 34: 123–137.
von Neumann J and Morgenstern O (1944) *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
Weymark JA (2022) Vaihinger's fictionalism meets Binmore's knowledge-as-commitment. *Homo Oeconomicus* 39. Forthcoming.
Wharton E (1904) The last asset. *Scribner's Magazine* 36: 150–168.
Wharton E (1991) *The Touchstone*. New York: HarperPerenniel. With a Prologue and Epilogue by Cynthia Griffin Wolff. Originally published in 1900.
Wiese H (2012) Backward induction in Indian animal tales. *International Journal of Hindu Studies* 16: 93–103.