

Sunk Costs, Rationality, and Acting for the Sake of the Past

THOMAS KELLY
University of Notre Dame

1. Introduction

Suppose that you are in the process of deliberating about how to spend the remainder of a given evening. You might attend a theater performance; alternatively, you might remain at home and read a novel. After careful reflection, you conclude that although you would enjoy either one of these activities, you would derive somewhat more enjoyment from remaining at home and reading the novel. In these circumstances, it would seem that remaining at home to read the novel is the rational choice, all else being equal.

Here, however, is a further fact about your circumstances. Some time ago, you purchased a ticket to this evening's theater performance at what you regard as a high price (say, \$200). Unfortunately, the ticket is non-refundable. Moreover, it is now too late to give the ticket away, or to resell it. If you do not attend tonight's theater performance, the expensive ticket will be *wasted*.

Would you be more likely to attend the theater performance, knowing that you had invested a significant sum in the acquisition of the ticket? (More likely than if you had been given the ticket for free, as the result of some promotional campaign?) Does the fact that you have invested a considerable sum in the ticket give you *a reason* to attend the theater performance, a reason that you would not possess if you had received the ticket for free? Given that you would in fact genuinely enjoy the theater performance, and remaining at home to read the novel would be only marginally more enjoyable, could your heavy past investment in the ticket actually tip the scales and make a difference to what it is rational for you to (decide to) do? Above, I suggested that if in fact you know that you would

derive more enjoyment from reading the novel than from attending the performance, then remaining at home to read the novel is the rational choice, all else being equal. Once it has been made clear that you have invested heavily in a ticket for the performance, is all else still equal?

If you would be more likely to attend the performance in virtue of having invested heavily in the ticket, then you tend to *honor sunk costs*. Many people—perhaps most people—do seem to give at least some weight to sunk costs in making decisions. There has always been considerable anecdotal evidence that this is so; in recent years, a number of psychological studies have been conducted which purport to provide experimental evidence for the same conclusion.¹

Questions pursued by psychologists about the extent to which individuals tend to honor sunk costs are descriptive questions. In addition to such (purely) descriptive questions, there are also a number of *normative* questions, or questions with normative implications, about the tendency to honor sunk costs. Is giving weight to sunk costs always irrational (as is typically assumed)? Would we be better off if we consistently ignored sunk costs? If we could eliminate the tendency to honor sunk costs, should we do so? Or are there circumstances in which giving weight to sunk costs has beneficial consequences and (hence) should be encouraged, or at least, not actively discouraged?

Although much of what I will say bears on the purely descriptive questions pursued by psychologists, my primary concern in the present paper is with normative questions of this sort. Such a concern might seem perverse. For it is widely agreed that honoring sunk costs is obviously and clearly irrational, and that doing so is, without exception, to be avoided. In economics and business textbooks, the tendency to honor sunk costs is treated as an elementary fallacy.² Psychologists who conduct experiments purporting to show that individuals do give some weight to sunk costs take these experiments as (yet further) psychological evidence of human irrationality and suggest possible remedies for this condition. It might seem then, that there are no interesting normative questions about the tendency to honor sunk costs, inasmuch as all such questions have easy answers. However, I will attempt to show that things are considerably less clear—and therefore, considerably more philosophically interesting—than economists and psychologists have sometimes assumed.³

The conventional wisdom about sunk costs, then, might be summarized as the conjunction of two claims:

- (1) Individuals often do give weight to sunk costs in their decision-making, and
- (2) It is irrational for them to do so.

The first of these claims encapsulates the conventional wisdom regarding the prevalence of the relevant practice; the second claim encapsulates the conventional wisdom regarding its normative status. A central aim of the

present paper is to undermine confidence in the conventional wisdom, understood as the conjunction of these two claims. (As we will see, *which* of the two conjuncts is the more dubious depends on subtle issues about how the phrase “give weight to sunk costs” is to be understood.)

Given its content, one who calls the conventional wisdom into question naturally appears as one who is concerned to defend the claim of human beings to being (largely) rational beings. My primary motivation for scrutinizing the conventional wisdom, however, is not rooted in a desire to defend human beings against the charge of irrationality. (Nor is it rooted in a desire to defend myself against the same charge.) Rather, my primary motivation is the following. The claim that it is irrational to give weight to sunk costs in one’s decision-making is naturally understood as a *constraint on the kinds of considerations that can legitimately be offered as, or taken to be, reasons*. And if in fact it is a common practice for individuals to treat sunk costs as reasons, then the constraint in question is an extremely substantive one. In the course of arguing that one possible action is superior to another, individuals often do seem to explicitly appeal to sunk costs: that is, they seem to treat sunk costs as though they are practical reasons. Such appeals are made in a wide variety of contexts, including public policy debates in which much is at stake. So, for example, politicians not infrequently demand further public funding of costly projects on the grounds that, if such funding is not forthcoming, previous expenditures will have been for naught.⁴ On occasion, the stakes may be even higher. Thus, in the final years of America’s military involvement in Vietnam, those who supported a policy of continued involvement frequently appealed to considerations of the following kind:

The United States has invested much in attempting to achieve its objectives. In addition to the many millions of dollars that have been spent, many thousands of lives have been lost, and an even greater number of lives have been irreparably damaged. If the United States withdraws from Vietnam without achieving its objectives, then all of these undeniably significant sacrifices would be wasted.⁵

There are, of course, numerous replies which might be made (and were made) to such arguments. It might be replied, for example, that one manifests a due appreciation for the past sacrifices of others not by continuing the policies that required such sacrifices, but rather by acting so as to eliminate (or minimize) the need for similar sacrifices in the future. However, if in fact it is always irrational to give weight to sunk costs, then it looks as though one need not waste time with such subtleties. For lives that are lost in the course of pursuing a given cause are sunk costs. American soldiers who have died in the course of pursuing the military objectives of the United States will be no less dead, whether the United States ultimately

achieves those objectives or simply abandons them. Because this is so, the fact that many lives have been lost in the past cannot provide a reason for favoring one course of action over another.

In short, the conventional wisdom about sunk costs seems to suggest an extremely substantive and quite general constraint on the kinds of consideration that can legitimately be offered as reasons for action. A primary concern of the present paper is to determine whether such a constraint is defensible, and, if it is defensible, exactly which arguments it would rule out as fallacious.

2. Some Conceptual Preliminaries

The tendency to honor sunk costs is easily confused with a number of other, distinct tendencies. It will be helpful then, to begin with some conceptual clarification.

Suppose that my psychology is such that, if I leave an expensive theater ticket unused, I will feel a considerable amount of regret at having done so. The psychological state of *feeling a considerable amount of regret* is, of course, a state which itself has a certain amount of disutility. In fact, its disutility might be such as to make a crucial difference as to whether I should attend the theater performance or stay home. That is, even, if staying home is in some sense “intrinsically preferable” to attending the theater performance—if I somehow forgot the actual circumstances in which I acquired the ticket, I would readily choose to stay home—it does not follow that it is rational for me to choose (1) over (2):

- (1) I stay home to read the novel and feel a considerable amount of regret at having wasted a \$200 ticket.
- (2) I attend the theater performance and feel no regret.

Notice that if I follow through on a certain course of action in order to avoid the regret that I would feel at having squandered previously-invested resources, my overt behavior will be identical to the behavior which I would manifest if I was honoring sunk costs. Moreover, in either case, the following counterfactual will be true:

If I had not previously invested heavily in this course of action, then I would not (continue to) perform it now.

However, performing an action in order to avoid future regret should be carefully distinguished from honoring sunk costs. In the former case, one treats the fact that *by performing this action, I will avoid a certain amount of future regret* as a reason for performing the action; in the latter case, one treats the fact that *I have already invested heavily in this action* as a reason

for performing it. To the extent that performing a certain action will affect my future psychological states in foreseeable ways, I should take these psychological consequences into account in deciding whether to perform the action, just as I should take into account any other foreseeable consequences. To do so is *not* to treat sunk costs as practical reasons.⁶

As we will see below, there are various other psychological tendencies which, although distinct from the tendency to honor sunk costs, manifest themselves in identical overt behavior across a wide range of situations. Failure to recognize that these other tendencies *are* distinct threatens to lead to mistakes with respect to both the normative and the descriptive issues surrounding sunk costs. So, for example, the fact that it might on occasion be rational to continue a given course of action in order to avoid future regret is irrelevant to the question of whether it is ever rational to honor sunk costs. With respect to the descriptive question, the fact that much behavior which might be explained as the manifestation of a tendency to honor sunk costs might also be explained as the manifestation of some distinct tendency clearly presents the danger that we will *overestimate* the pervasiveness of the tendency to honor sunk costs. In fact, much of the psychological literature which purports to demonstrate the pervasiveness of the tendency to honor sunk costs is, I believe, methodologically flawed in exactly this way.

Perhaps the clearest case is a widely-cited study by Hal Arkes and Catherine Blumer entitled “The Psychology of Sunk Cost”. After reporting on ten independent experiments, each of which, the authors claim, supports the view that there is a widespread tendency to honor sunk costs, the authors propose an explanatory hypothesis in order to account for this behavior. Their explanation runs as follows: individuals often continue to pour resources into already-begun projects because they fear that, if they abandon a project in which they have already invested heavily, they will be perceived as being *wasteful* by others (p. 132). Now, if this explanatory hypothesis is true, then it undermines the authors’ claim that the behavior in question is indicative of a tendency to honor sunk costs. That is, the hypothesis that people continue already-begun projects in order to avoid appearing wasteful *competes* with the hypothesis that the same behavior is explained by their treating sunk costs as reasons for action. *Not being perceived as wasteful by others* is, after all, a potentially valuable state of affairs. Insofar as I am following through on a given course of action in order to bring about this potentially valuable state of affairs, I am no more giving weight to sunk costs than I am when I act in order to bring about any other worthwhile end. In such circumstances, it is true that, if I had not previously invested in a certain course of action, then I would have no reason to continue now: that I have a reason now is historically or counterfactually dependent on my having made this past investment. However, my reason for continuing is not provided by my having made the past invest-

ment itself; rather, my reason is provided by the fact that *by performing this action, I will be able to avoid being thought wasteful by others.*⁷

The issue of how the tendency to honor sunk costs is to be demarcated from other, distinct tendencies is a crucial one, and an issue to which we shall return below.

3. Honoring Sunk Costs: Some Possible Advantages

It is widely assumed that rationality requires that one ignore sunk costs in one's decision-making. What underlies this assumption? Here is a natural thought: rationality requires that one ignore sunk costs because those who scrupulously ignore sunk costs *fare better* than those who do not. Firms that give some weight to sunk costs tend to go out of business more frequently than firms that ignore sunk costs; individuals who give some weight to sunk costs fare worse with respect to achieving their goals than individuals who differ from them in this respect.

The claim that individuals who give some weight to sunk costs fare worse is, however, an empirical claim, and one which is false if certain not implausible assumptions are true. Suppose, for example, that there is a significant statistical correlation between the tendency to give some weight to sunk costs and the character trait of *being strongly averse to waste*. (Plausibly: individuals who are strongly averse to waste are, as a class, more likely to be influenced by the knowledge that they have already spent \$200 on a ticket than individuals who have no such aversion and remain unmoved in the face of fruitless expenditures.) And the character trait of being strongly averse to waste might in turn give rise to greater efficiency on the whole. That is, the tendency to honor sunk costs might be positively correlated with some other character trait, a trait whose net effect is to promote the goals of those who possess it. Thus, it is far from obvious that those who consistently ignore sunk costs fare better than those who do not.

However, even if those who consistently ignore sunk costs fare no better as a class, we might still expect that those who consistently ignore sunk costs *and* have a strong aversion to waste fare best of all. And isn't this narrower class the relevant comparison class? Indeed, perhaps a deliberate policy of always ignoring sunk costs is the *theoretically sophisticated* manifestation of an aversion to wasteful behavior. We still have no reason to believe that honoring sunk costs can itself prove advantageous—even if certain traits with which it is correlated are advantageous. Are there circumstances in which the tendency is advantageous?

Robert Nozick has suggested that our tendency to honor sunk costs plays a beneficial role in enabling us to overcome temptation.⁸ Because we are reluctant to abandon projects in which we have invested large amounts of resources, we are less likely to abandon those projects in the face of temptation, and this tends to promote our long-run well-being. (Here is an example to

illustrate Nozick's idea. An individual is strongly tempted to cheat on his spouse. If he does so, he will later regard it as a mistake for the rest of his life. The individual ultimately refrains, in part because he reflects on how much he has invested in his marriage—financially, emotionally, and temporally. These investments are, of course, sunk costs. But because he allows such allegedly irrelevant considerations to influence his decision-making, he is ultimately better off than he would be if had ignored them and succumbed to temptation.⁹)

We might wonder whether the tendency to honor sunk costs can prove advantageous in the absence of a tendency to succumb to temptation, or some other character defect. Granted that the tendency to honor sunk costs might prove beneficial in offsetting weakness in the face of temptation, would agents who are not flawed in the relevant respects also benefit in foreseeable ways from this tendency?

In fact, there is a class of game-theoretic situations in which an agent who honors sunk costs *is* better off, even if he or she does not suffer from weakness of the will. Because these situations have a certain practical importance as well as theoretical interest, they are worth examining at some length.

Recall our earlier example of the Vietnam War. In the Vietnam War, the United States faced a Communist adversary of manifest inferiority. There was no real possibility of the Communists ever destroying (or indeed, significantly impairing) the ability of the United States to continue to pursue its objectives in Indochina. Moreover, this fact was well-understood by both sides. Nevertheless, the Communists ultimately achieved all of their objectives, and the Americans manifestly failed to achieve any of their own. How was this possible?

Here is a short but true answer to this question: the Communists succeeded in making it clear to the United States that continued pursuit of its objectives would result in significant costs, costs that were ultimately judged to be not worth bearing. Often, a relatively weak party will achieve its goals in the face of the open antagonism of a much stronger opponent by convincing the much stronger opponent that the price of continued struggle is simply not worth paying. Because this is so, a weaker opponent will have every reason to maximize the costs that are incurred by its stronger opponent as the stronger opponent pursues its goals—these costs then serve as evidence that further pursuit of the relevant goals will result in similar costs in the future. This, of course, is the principle by which much terrorism operates: terrorists rarely have any intention, nor any realistic hope, of significantly impairing the ability of the large states which they terrorize to continue their policies. Rather, the point is to convince the large state that continued pursuit of those policies will carry with it unacceptably-high costs.

Notice, however, that this (often quite effective) terrorist strategy presupposes that there are very real limits as to how much weight the stronger opponent gives to sunk costs. In order to more fully appreciate this point, it will be helpful to contrast two ideal types: the expected utility maximizer and the **pure honorer of sunk costs**. The expected utility maximizer is simply the ideally rational agent as portrayed by classical economics. As such, she scrupulously ignores sunk costs in her decision-making. On the other hand, the pure honorer of sunk costs deliberates in the following manner: the more resources he has invested in pursuing a given course of action in the past, the stronger the reasons he takes himself to have to continue that course of action now and in the future (so long as there is any hope of achieving the relevant goals at all). Thus, the nation which decided its foreign policies as a pure honorer of sunk costs would have the following policy towards terrorists: the more of our people who have been victimized by terrorist activities in the past, the more determined we are to stay the course.

Notice that while the typical terrorist tactic of maximizing inflicted costs is an effective strategy against the expected utility maximizer, it is utterly counterproductive against the pure honorer of sunk costs. For the more children the terrorists manage to blow up with bombs, the more determined the victimized nation is to continue those policies that originally prompted the bombings, a process that in effect thwarts the desires of the terrorists.

Such a stubborn policy of honoring sunk costs might seem to be a disastrous policy for the terrorized nation itself, one which effectively condemns many of its future children to death *via* terrorist activities. But this natural thought ignores the reputation effects that might accompany being a pure sunk cost honorer. For over time, would-be terrorists will see that each “successful” terrorist attack only hardens the resolve of the victimized nation to carry on with its despised policies. At some point, the terrorists will realize: *these people honor sunk costs*. At that moment, the would-be terrorists acquire strong reasons to discontinue the policy of maximizing the costs incurred by the larger nation. In contrast, the pure expected utility maximizer is at the mercy of the technological wherewithal of the terrorist, for the terrorist knows that he can successfully manipulate the behavior of his prey if only he can solve the practical problem of finding a way to make the costs high enough.¹⁰

Since the pioneering work of Thomas Schelling, it has been well-known that there are game-theoretic situations in which individuals and nations benefit in foreseeable ways in virtue of being less than perfectly rational.¹¹ Thus, nations that are practically irrational do not get threatened as often, because would-be threateners cannot count on the irrational nation to respond in the desired way. In being (known to be) rational, one incurs a certain vulnerability to being manipulated *via* threats. Our present point is considerably more radical. It is not only that there is no reason for a weaker party to maximize the costs incurred by the pure honorer of sunk costs.

Rather, in such circumstances, the weaker party has strong reasons to take active measures to *minimize* the costs that the more powerful nation incurs in pursuing its detested objectives. If the typical terrorist tactic is *ineffective* against one who is not rational, it is positively *counterproductive* against one who honors sunk costs. If the United States was known to be a pure honorer of sunk costs, then the best strategy for the Communists in Vietnam might very well have been to *subsidize* a continued American military presence in Vietnam. For the Communists would know that, the more financial resources that the Americans were forced to expend, the more reluctant the Americans would be to leave. The Communists' best hope might have been to wait until changing geopolitical circumstances obviated the perceived benefits of an American presence in Vietnam.

(Here is a somewhat more realistic example to illustrate the same point. A low-level member of a terrorist organization, enraged by the policies of a hated nation, is motivated by revenge to launch a terrorist strike. The leader of the terrorist organization, who knows that the hated nation honors sunk costs, realizes that such an attack will have only counterproductive consequences with respect to the ultimate achievement of the terrorists' goals. The leader thus has strong reasons to do everything that he can to restrain the vengeful impulses of his followers, and the nation that honors sunk costs benefits from this.)

There are then, possible situations in which the pure honorer of sunk costs fares better than the pure expected utility maximizer.¹² To say this, of course, is not to say that we would be better off, either individually or collectively, as honorers of sunk costs than as expected utility maximizers. Whether it is advantageous to give (some level of) weight to sunk costs depends on how often contexts arise in which it would be advantageous to give (that level of) weight to sunk costs, how advantageous doing so is in those contexts, as well as on parallel facts about all of the many contexts in which it would be *disadvantageous* to do so. Best of all, of course, would be to give weight to sunk costs in all and only those contexts in which doing so is advantageous, but dispositions are seldom so agreeably fine-tuned in their manifestations.

Before moving on, I would like to record a certain pessimistic speculation. Often, individuals and nations, being neither pure expected utility maximizers nor pure honorers of sunk costs, suffer greatly from giving some weight to both in their decision-making. Again, the Vietnam example is instructive. Because we as Americans give some weight to sunk costs in deciding upon a policy (some prominent political leaders, and many ordinary citizens, feel that "we need to stay in, because we've invested so much already"), we stay in considerably longer than we would have otherwise and thus incur much greater costs than if we were pure expected utility maximizers. On the other hand, because we give a great deal of weight to expected utility (there are very strong limits as to how much weight we

give to sunk costs), and the Communists know this, we receive none of the benefits which we might enjoy from the reputation-effects of being sunk cost honorers. Thus, the particular mixture of weight given to considerations of expected utility and to sunk costs might result in peculiarly bad policies.¹³

Suppose that I know that I would be better off, on the whole, if I gave some weight to sunk costs than if I consistently ignored them. (Perhaps I am peculiarly prone to succumbing to temptation and a tendency to honor sunk costs would help to offset this destructive habit, or perhaps I stand to benefit from a reputation of steadfastly pursuing already-begun projects.) What difference would this knowledge make to what it is rational for me to do? If I do not already have the advantageous disposition, but there is some way in which I might acquire it, then (all else being equal) it will be rational for me to act so as to acquire it. Alternatively—and perhaps more realistically—if I am already in the habit of giving weight to sunk costs, then (all else being equal) it would be irrational for me to attempt to rid myself of this habit.¹⁴ Thus, the fact that honoring sunk costs can prove foreseeably advantageous might make a difference to what it is rational to do.

Imagine next someone who has overriding reasons to acquire the disposition and acquires the disposition for those reasons. When he subsequently manifests the disposition and benefits as a result, is he rational in so manifesting it?

This question is controversial. Some theorists, notably David Gauthier, would answer it in the affirmative. According to Gauthier, “a disposition is rational if and only if having it is most conducive to one’s substantive aim” (1997: p. 31). Others disagree. Derek Parfit, for example, has argued that one can be irrational in manifesting a disposition which best promotes one’s aims, even if one rationally acquired that disposition because one realized its advantageousness (1984, ch.1). For Gauthier, “this conclusion should make us suspect Parfit’s account of rationality” (1997, p. 31).

Notice that if Gauthier is correct, then there will be (possible) individuals who take sunk costs into account and are rational in doing so. For there are possible individuals who would best achieve their substantive aims if they gave some weight to sunk costs in their decision-making. On the other hand, if Parfit is correct, then the mere fact that there are possible individuals who would best achieve their substantive aims by honoring sunk costs does nothing to show that honoring sunk costs can be rational. Although I believe that Parfit has the stronger case with respect to this particular issue, I have little to add to the reasons which Parfit and others have offered in defense of their view—reasons which Gauthier evidently does not find compelling.¹⁵

Fortunately, we need not settle this issue here. Having noted that *if* Gauthier’s view is correct, then there are possible cases in which honoring sunk costs is rational, I will simply proceed, in the remainder of the paper, as though Gauthier’s view is false. This will enable us to examine the

normative status of the tendency even if “being most conducive to one’s substantive aim” is not in fact a sufficient condition for its rationality.

4. Sunk Costs and the Concept of a Redemptive Action

The claim that it is characteristic of rational decision-making to ignore sunk costs is naturally cast as the claim that *ignorance* of sunk costs is no handicap with respect to decision-making. Famously, John Rawls held that the subject of justice could be illuminated by considering the deliberations of self-interested agents situated behind a veil of ignorance, a veil which effectively shields them from information which (Rawls held) could only have a distorting influence on their judgements (1972, especially pp. 17–22). We might attempt to mimic Rawls’ procedure and illuminate the subject of sunk costs by considering the position of agents situated behind another veil of ignorance, a veil which effectively shields them from information about past investments in various courses of action.

Let us make the thought experiment more concrete. Imagine that we possessed a pill which had the following peculiar property. Upon taking the pill, an agent who is faced with a choice between alternative actions $A_1 \dots A_n$ immediately forgets any past investments (either on her own part or on the part of others) in any of the alternatives under consideration. This state of forgetfulness persists for the duration of her deliberations; once a decision has been made, the relevant memories immediately return. Here is a (rough) suggestion for making vivid the doctrine that rationality requires that sunk costs be ignored. The theorist who counsels that sunk costs should be ignored might insist that an agent who decides upon a course of action while under the influence of the pill is in *at least as good* of a position to make the correct decision as she would be if she were not under its influence. Indeed, such a theorist might insist that the agent is in a *better* position to make the correct decision while under the influence of the pill, inasmuch as she does not run the risk of being unduly influenced by irrelevant factors. The perfectly rational decision maker—one who, among other things, consistently gives no weight to sunk costs—would make the same decision regardless of whether she was on the pill. Alternatively: the perfectly rational decision maker who makes a decision while under the influence of the pill would never *regret* her decision, simply because the relevant memories have returned.

However, the doctrine that sunk costs should be ignored is not perfectly captured by the claim that there is no disadvantage to pill-influenced decision-making. The theorist who claims that rationality requires the ignoring of sunk costs might consistently agree that, in certain circumstances, choosing behind the envisaged veil of ignorance is disadvantageous. What are those circumstances?

(1) In some cases, the fact that I have previously invested in a given course of action might constitute evidence that that course of action is likely to succeed or fail. Suppose that, in anticipation of becoming a medical doctor, I have previously invested heavily in training, education, and so on. If, at some point, I consider whether to continue my pursuit of this goal, I might underestimate the chances of ultimately achieving the goal if I do not take into account how much I have already invested in its pursuit.

Alternatively, the fact that I have invested heavily in a given course of action might constitute evidence that that course of action is *unlikely* to succeed. Suppose that the United States has already invested much in attempting to defeat the Communists in Vietnam but has not yet succeeded. If these past investments are ignored, then Americans might overestimate the chances of ultimately achieving this goal—or else, underestimate the future investments which will be required in order to achieve that goal.

In cases in which past investments in a given course of action influence the probability that that course of action will be successful if continued, a decision-maker deliberating behind a veil of ignorance will be at a disadvantage. This, of course, will be readily conceded by the defender of the view that sunk costs should be ignored: his view is that sunk costs should be ignored *qua* sunk costs, not that they are *epistemically* irrelevant.

(2) As we have already noted, in some cases the future value of successfully carrying out a course of action might depend on whether one has made a significant past investment in that course of action. Thus, if others will think that I am wasteful if I fail to achieve some goal in whose pursuit I have already invested considerable resources, then the value of actually achieving any particular goal will depend on whether or not I have already invested heavily in its achievement. This too, will readily be conceded by the theorist who counsels that sunk costs should be ignored.

Is the value of knowing about sunk costs *exhausted* by the relevance that such knowledge has to deciding cases of type (1) and (2)? Notice that, insofar as the value of such knowledge is exhausted by its relevance to these two kinds of case, that value is highly derivative, in the following sense. What is really valuable for present decision-making is being able to predict the future consequences of possible actions; knowledge of sunk costs is valuable only insofar as such knowledge enhances one's ability to predict those future consequences. Conversely: ignorance of sunk costs is no handicap to present decision-making (and indeed, might very well be of some benefit) so long as that ignorance does not hinder one's ability to predict the future.

The essentially derivative importance that an awareness of sunk costs has on this picture can be brought out in the following way. Let us stipulate that the decision problem facing an agent behind our envisaged veil of ignorance is a **decision under certainty with respect to the future**, in the sense that the agent knows that, if he performs act A_1 , then the future history of the world

will be H_1 , while if he performs the alternative action A_2 , the future history of the world will be H_2 . (Suppose, for the sake of simplicity, that there are only two alternatives under active consideration. We may also suppose that in neither of these future histories will the agent violate any deontological constraints.) These future histories are *complete*, in the sense that they include information about any future regret that the agent will feel, the reputation that he will enjoy among others, and so on. It is tempting to conclude that an agent with such a perfectly clear view of the future has all of the information that he could ever need or want, and that his ignorance of sunk costs is of no handicap. For shouldn't he simply decide which future history of the world is better, and then act accordingly? Full knowledge of the future consequences of pursuing a course of action seems to obviate the need to know of any past investments in that course of action.

This intuitively appealing line of thought is, I believe, mistaken. In some cases, a knowledge of sunk costs is *essential* to knowing what one should do. An agent who lacked such knowledge would be at a disadvantage with respect to his decision-making, even if he had perfect knowledge of the future consequences of his actions.

Before arguing for this conclusion, I want to examine the source of the intuitive appeal of the contrary view. Why is it natural to suppose that while the future consequences of performing an action are highly relevant to the question of whether that action should be performed, past investments in that course of action are wholly irrelevant? The following, I think, is a natural answer to this question:

While actions that we perform now can affect the future, they are powerless to affect the past. Sunk costs are, by definition, costs that cannot be recovered, no matter how the agent decides to act. And it is because of this that they can and should be ignored.

However, the claim that present actions cannot affect past events requires a certain amount of interpretation. It is a philosophical commonplace that the significance of past events might depend on the causal relations which they bear to future events, future events whose occurrence (or nonoccurrence) might depend on what we do now. A criminal seriously wounds an innocent person; hours later, a doctor struggles to save the victim's life. Whether the criminal has or has not committed a murder depends on whether the doctor's later efforts are successful. Some years ago, Arthur Danto emphasized the extent to which many of the most significant descriptions of prominent historical events depend for their applicability on the subsequent (and often highly-contingent) occurrence of later events (1985). More recently, a number of writers have held that the significance of particular events within an individual life can depend on how they are embedded within larger narrative structures; the value of events which are identical with respect to their intrinsic properties

might vary greatly depending on how they are embedded within larger contexts. Consider, for example, David Velleman on how *learning from a misfortune* can help to transform the significance of an earlier event:

The point of learning from a misfortune, surely, is to prevent that misfortune from being a total loss. Learning from the misfortune confers some value on it, by making it the means to one's edification . . . conferring instrumental value on a misfortune alters its meaning, its significance in the story of one's life . . . an edifying misfortune is not just offset but redeemed, by being given a meaningful place in one's progress through life (2000, p. 65).

More generally

. . . later events . . . alter the significance of earlier events, thereby altering their contribution to the value of one's life (2000, p. 65).

Charles Taylor has similarly written of our desire for "the future to 'redeem' the past, to make it part of a life story which has sense or purpose" (1989, p. 43).¹⁶

A common theme of such thinkers is this: present actions can affect the significance of past events in ways that matter deeply to us. This fact, I think, has important implications for the topic of sunk costs. For it is because present actions can affect the significance of past events in ways that matter deeply to us that an individual who had a perfectly clear view of the future, but who was wholly ignorant of sunk costs, would be at a distinct disadvantage with respect to his ability to satisfy certain quite natural preferences.¹⁷ That is, knowledge of sunk costs is *essential* to the fulfillment of certain preferences. Let us describe one such preference. One might prefer that, if others have made significant sacrifices in attempting to realize some valuable state of affairs S, then their sacrifices *not be in vain*. That is, one might prefer that these sacrifices causally contribute to the realization of some valuable state of affairs. (Perhaps one regards it as important that the sacrifices in question result in the achievement of the originally-intended valuable state of affairs S; or perhaps one thinks that what is important is simply that *some* valuable state of affairs results. These are, of course, importantly different preferences, although satisfaction of the first suffices for satisfaction of the second.) Interestingly, one sometimes *is* in a position to determine, by one's own actions, whether the past efforts of others will have been in vain. This is true, for example, when it is within one's power to finish some valuable project in whose service others have labored, but which they are now not in a position to complete.

Let us say that when one acts so as to prevent the past efforts of others from having been in vain one **redeems** those efforts. One might regard it as valuable that the efforts of others not be in vain, and thus, regard the fact

that a given action is a redemptive action as an (additional) reason for performing it. Notice that if I do have such a preference for redemptive actions, then being ignorant of sunk costs will effectively frustrate my ability to satisfy it. For whether or not a given action is the action which I have most reason to perform might depend on whether that action is a redemptive action. And whether that action is a redemptive action will depend in turn on whether there is some valuable state of affairs *S* which would be both (i) a causal consequence of my action, and (ii) a causal consequence of the efforts of others, efforts which would otherwise remain unredeemed.¹⁸

In short: the fact that significant investments have been made in the course of pursuing some goal *G* can give rise to new reasons to pursue *G* (in addition to whatever reasons originally existed for pursuing *G*). And in a given case, these new reasons might make a decisive difference to whether one should act so as to bring about *G*, or in some alternative way instead.¹⁹ It is because such cases are possible that knowledge of sunk costs is sometimes essential to knowing what one should do.

There is nothing, I believe, irrational about such “backward-looking” preferences. This claim will be readily conceded by those who hold that preferences are not subject to rational criticism, as well as by those who hold, more weakly, that although sets of preferences can be jointly irrational, no particular preference can be intrinsically irrational. But even if we adopt a less liberal view, and admit that some preferences are intrinsically irrational,²⁰ it is unclear why preferences for redemption should be included among this class. (We will return to this issue briefly in Section 6 below.)

If we take seriously the idea that the significance of past events can be altered by what happens in the future, then we should be open to the following possibility. An agent behind the veil of ignorance decides that future history H_1 is better than future H_2 , and chooses accordingly. If, however, he was not ignorant of sunk costs, he would choose to act so as to actualize H_2 , for the redemptive power of H_2 significantly exceeds that of H_1 .

The possibility that one might have an independent preference which motivates one to pursue old goals is noted in passing by Steele (1996, p. 617). He claims—correctly, I believe—that there need be nothing irrational about such preferences, or about attempting to fulfill them. However, he regards the preferences which generate such reasons as eccentric: in the space of a paragraph, he refers to the circumstances which give rise to them as “bizarre”, “unusual” and “peculiar”. This, I think, is a mistake. On the contrary, such preferences are quite common. Above, we noted that one might have a preference that the past sacrifices of others not be in vain. Although this preference is, I think, fairly common, it is perhaps much less common than its self-centered analogue, viz. that one’s *own* past efforts and sacrifices not be in vain.

The claim that individuals often strongly prefer that their own past efforts be redeemed receives support from an interesting phenomenon which has been emphasized by Gilbert Harman. Harman observes that, so strong is our desire to see our own past efforts play a role in bringing about valuable ends, we will often adopt *new* ends, carefully tailored, so that our past efforts can be seen as instrumentally valuable means to the achievement of these ends.

Sometimes one adopts an end so that things that one has already done . . . can gain significance as a means toward the end that one now adopts. This is how some people choose careers—so that their earlier training will not be wasted.

The typical teacher of physics, for example, did not originally study the subject with the purpose of becoming a physics teacher. He studied the subject because it interested him or perhaps because it was a required subject. Later he decided to become a teacher of physics; at that point he was able to see his earlier study as part of the means by which he is able to become a physics teacher—and that is part of his reason for becoming a physics teacher: so that his earlier study will not have been wasted. It is true that he enjoys physics (let us suppose), so he is choosing a career that will allow him to do something that he enjoys doing. But that is only one consideration. It is also relevant that that in becoming a physics teacher he gives significance to his earlier study and to that extent helps to unify his life. If he had chosen instead to do something else . . . his training in physics might . . . have had no significance—it would have been wasted. It would have lost connection with his later life, a connection that could have helped to give his life unity . . . Past acts done for other reasons assume a pattern in the light of ends adopted only now.

We adopt ends that help to rationalize and give significance to what we have been and are doing—not only in large decisions, as in choosing careers, but even in our smallest and most insignificant acts (1999, pp. 72–73).

Here, Harman is concerned with the adoption of new ends as a way of bestowing meaning and significance upon one's previous efforts. This differs from the phenomenon with which we have been primarily concerned, in which an agent continues to pursue an old end so that her previous efforts on its behalf will not be in vain. In both cases, however, the underlying motivation is the same: the desire to see one's previous efforts redeemed. We should expect that, often, of the various ends which one might pursue in order to redeem some previous effort, it will be most rational to pursue that end E in whose pursuit the effort was originally expended. After all, because the effort was originally put forth with E in view, the particular actions in which it manifested itself were selected because of their perceived efficacy as means for achieving E, as opposed to any other end. We would expect then, that the desire that one's previous efforts not be in vain would often lead to the continued pursuit of the end towards which the efforts were originally directed.

5. The Role of Past Sacrifices in Practical Deliberation: Some Proposals

I have suggested that an agent who treats the fact that a given action would be a redemptive action as a reason for performing it need not be irrational in doing so. One who makes such a suggestion naturally invites a host of further questions. How exactly should considerations of the past fit into one's all-things-considered deliberations? Are there normative principles which dictate when one ought to write off sunk costs, notwithstanding that one will have to chalk them up as a waste? What (if anything) underwrites the value of redemption, or the value of having a life which possesses narrative unity?

I believe that answering these and related questions in a wholly satisfactory way would require something approaching a full-fledged theory of practical reasoning, a theory which I do not possess. Nevertheless, in this section I offer some proposals which bear on the relevant issues.

Let us begin by noting some limitations on the capacity of past sacrifices to generate present reasons for action. Above, I claimed that the fact that significant sacrifices have been made in the course of pursuing some goal *G* can give rise to new reasons to pursue *G*, and that, in a given case, these new reasons might make a decisive difference to whether one should act so as to bring about *G* or in some alternative way instead. In attempting to make these claims plausible, I focused on cases in which there is originally some independent reason for pursuing *G*, apart from any past sacrifices that have been made in its pursuit. A more radical view would hold that, even if there were originally no reason for pursuing *G*, the fact that in the past much effort has been (mistakenly) spent in its pursuit can now provide a genuine reason for continuing to pursue it. If this more radical view were true, then goals which there was originally no reason to pursue might “bootstrap” their way into being rationally pursued, simply in virtue of having been pursued in the past.²¹ But this more radical view should, I think, be rejected. Its inadequacy is perhaps clearest in cases in which the original pursuit of the goal *G* is motivated by false beliefs about the value of its achievement. Thus, suppose that I have expended considerable energy and effort in an attempt to locate a particular book because I believe that it contains the answer to some question about which I am curious. In virtue of these past efforts, I have managed to narrow down the range of the book's possible locations; I am thus closer to finding it than I was at the outset of my search, and, if I now proceed to find it, these past efforts will have played an essential role in the causal history of my doing so. Suppose, however, that before I am able to find the book in question, I learn that it does not in fact contain the answer to my question. In these circumstances, it would be irrational, I believe, to continue my search simply because otherwise my past efforts will have been for naught. Here then is a substantive constraint on the potential of past sacrifices to generate present

reasons for action: past sacrifices expended in pursuit of a goal G can generate further reasons for pursuing G only if there is some independent reason to achieve G.

Another case in which it seems that past efforts which have yet to bear fruit can simply be written off without risk of offending against some significant value is when those efforts are constitutive parts of some larger effort or project that has itself proven fruitful. Notice that not every constitutive part of a fruitful project need itself be fruitful, or contribute to the fruitfulness of the whole. Thus, imagine a graduate student who has recently completed a successful dissertation. During the years which she has devoted to this project, some of the lines of inquiry which she has pursued have not resulted in significant contributions to the finished version. (Indeed, we can suppose that the ultimate content of the dissertation is not in any way affected by the fact that these lines of inquiry were pursued.) Nevertheless, from her present perspective, the graduate student views these efforts as constitutive parts of the larger, temporally extended project which ultimately culminated in the successful dissertation. The fact that these fruitless efforts are naturally viewed as part of a larger project that did prove fruitful seems to obviate any need to redeem them (despite the fact that, strictly speaking, they did not causally contribute to the value manifested by the finished product).²²

What, if anything, accounts for the value of redemption in those cases in which it *is* plausibly taken to be of some value? The fact that in some circumstances we (or at least, many of us) are inclined to value redemption seems to derive from a fundamental respect for persons and their projects. It is not simply the fact that precious resources have been expended but rather that the relevant costs have been incurred by agents pursuing particular projects which we take to be worthwhile. After all, a strictly natural process might consume large amounts of valuable resources (redwoods trees, or fossil fuels). As a result of such a process running its course, we might find that some valuable state of affairs has been brought within “striking distance”: it now lies within our power to achieve this state of affairs, and, if we do so, the past consumption of these valuable resources will have played an indispensable role in the causal history of its achievement. Whether we ought to devote further resources to achieving the relevant state of affairs depends, I think, solely on how valuable its achievement would be compared to what else we might achieve by using these further resources in alternative ways: the mere fact that (say) significant numbers of redwoods have already been consumed is not, I think, a consideration which should enter into our deliberations at this point. In contrast, the fact that one’s ancestors have toiled for generations in the pursuit of some goal (a goal which one recognizes as having some positive value) can, I think, properly enter into one’s deliberations as a consideration which counts in favor of devoting further resources to its pursuit. In this way, the value of

redemption seems to derive from a more fundamental respect for agents and their projects.

If this speculation about the source of the value of redemption is correct, then it naturally suggests another general principle as to how we should and should not attempt to redeem the past sacrifices of others. In the last section, I noted that we should distinguish between two importantly different preferences that one might have:

- (i) that the past sacrifices of others causally contribute to the realization of some valuable state of affairs, and
- (ii) that the past sacrifices of others causally contribute to the realization of that valuable state of affairs in the pursuit of which those sacrifices were originally made.

If in fact the value of redemption derives from a more fundamental respect for agents and their projects, then the second preference might seem to constitute a more fitting manifestation of such respect (at least, all else being equal). Similarly: even if one does attempt to redeem the past sacrifices of others by bringing about some state of affairs that they would not have recognized as valuable, it seems important that the individuals in question would at least not have despised the relevant outcome, or have been horrified by the knowledge that their past efforts played a role in the causal history of its realization.

We have emphasized that one who values redemption will tend to prefer redemptive actions to non-redemptive actions. It is worth noting though, that one who values redemption will allow this fact to enter into her practical deliberations in other ways as well. Thus, valuing redemption might give one reasons to choose some redemptive actions over other redemptive actions: even if A_1 and A_2 are both redemptive actions, A_1 might redeem all of the past efforts which A_2 redeems, and some additional past efforts as well. Or perhaps, more individuals would have their efforts redeemed by A_1 , or the sacrifices which A_1 redeems might be more significant. That is, some redemptive actions might be more redemptive, or have greater redemptive power, than others. Conversely, an agent who values redemption might have reasons for preferring some non-redemptive actions to other non-redemptive actions: even if doing A_1 does not suffice to redeem certain past efforts, it might bring those efforts closer to redemption than doing A_2 . (Because A_1 has been performed, the probability of the efforts ultimately being redeemed is increased, or it is now easier for others to redeem the past efforts than it would be if A_2 had been performed.) Or perhaps, even if doing A_1 does not raise the probability of redemption, it does not lower that probability, which would be a result of doing A_2 . Clearly, there are many possibilities here.

6. Conclusion: The Conventional Wisdom Reconsidered

Suppose that I have a preference that my past efforts not be in vain. Suppose further that because I have this preference, I sometimes perform actions which are intended to further projects in whose service I have made significant efforts in the past, actions which I would not perform were it not for their status as redemptive actions. This preference is not, I have claimed, irrational. And if this preference is not irrational, the behavior which I exhibit in attempting to better satisfy it need not be irrational either.

If my behavior is not irrational, then the salient question becomes: in behaving in this way, am I honoring sunk costs? For if my behavior does count as a genuine case of honoring sunk costs, and my behavior is not irrational, then we can immediately conclude: giving weight to sunk costs is not always irrational.

Whether my behavior does count as giving weight to sunk costs depends on what is meant by “giving weight to sunk costs”. In discussions of sunk costs, some participants use the relevant terminology in an inclusive manner. Thus, Arkes and Blumer define “the sunk cost effect” as “a greater tendency to continue an endeavor once an investment in time, effort, or money has been made” (1985, p. 124). Given such a wide definition, my behavior will count as a genuine case of giving weight to sunk costs. Such wide definitions are, unsurprisingly, popular among psychologists seeking to show that human beings often do give weight to sunk costs in their decision-making.

On the other hand, others who have discussed sunk costs have circumscribed the relevant tendency more narrowly. As we have noted, Steele explicitly distinguishes the honoring of sunk costs from cases in which one has an independent preference for finishing already-begun projects. On such a narrow construal, my behavior will not qualify as a genuine case of giving weight to sunk costs, and so the rationality of my behavior does nothing to show that giving weight to sunk costs is (ever) rational.

There are, I believe, reasons for preferring the less inclusive terminology.²³ However, it is not my intention to (attempt to) legislate how the relevant terminology should be used. Rather, I want to suggest that the fact that different theorists have used the same terminology in different ways should have important consequences for our confidence in the conventional wisdom about sunk costs. In Section 1, I claimed that that conventional wisdom could be summarized as the conjunction of two claims:

- (1) Individuals often do give weight to sunk costs in their decision-making, and
- (2) It is irrational for them to do so.

My present suggestion is that we should be much less confident that the conventional wisdom is true, once we recognize that the phrase “giving

weight to sunk costs” is highly susceptible to being used equivocally. Suppose, for example, that we employ a *wide* definition of what it is to give weight to sunk costs—according to which (say) one is giving weight to sunk costs insofar as past investments in a given course of action make it more likely that one will choose to continue that course of action. In that case, we can expect that there will be no shortage of evidence—either anecdotal or the kind of evidence supplied by the psychology laboratory—that we *do* honor sunk costs. Given such a wide definition, however, the behavior in question need not be irrational. That is, if we interpret the phrase “giving weight to sunk costs” in such a way as to make claim (1) clearly true, then the normative claim (2) will not be compelling. On the other hand, if we employ a narrow definition, then the normative claim will be (much more) defensible.²⁴ Given such a narrow definition, however, it then becomes unclear whether giving weight to sunk costs is something which people actually do with any interesting degree of frequency. That is, given a narrow definition, (2) will be defensible, but (1) becomes dubious. Indeed, given such a narrow definition, many of the paradigm examples which are often used to illustrate the tendency do not count as genuine cases of honoring sunk costs at all. Consider, once more, an example which is often employed as a paradigm case of giving weight to sunk costs, and indeed, one which we ourselves have employed as such throughout the present paper: the Vietnam War. Those who appealed to lives already lost as reasons for staying the course in Vietnam are naturally understood, I believe, as advocating such a policy as a means of *redeeming* those lives. Again, we might criticize such appeals: on the grounds that the value of redemption should not be given priority over other, competing values (e.g., the value of saving future lives), or perhaps, on the grounds that continued prosecution of a war which many believed to be deeply immoral *could not* result in the redemption of those lives. To criticize such an appeal in either of these ways, however, is not to criticize it on the grounds that it embodies a sunk cost fallacy.

We might wonder: given a narrow definition, what *would* count as a genuine case of giving weight to sunk costs? Consider two different considerations which might be cited as reasons for continuing some war:

- (1) We have already lost many lives in this war.
- (2) We have already lost many lives in this war, and if we pull out now, these lives will have been wasted. If, on the other hand, we continue to prosecute this war, these lives might be redeemed.

Imagine someone who claimed that (1) provided a reason for staying the course, but who truly denied that this consideration could be more accurately captured by (2), or by some similar formulation. That is, imagine someone who takes the *mere* fact that much has been lost in pursuing some goal as a reason for continuing to pursue that goal. This, I think, would be a

genuine case of giving weight to sunk costs, even on a narrow definition. Such individuals are imaginable, but they are not, I suspect, often encountered in real life. It would be unwise, of course, to assume that someone who explicitly cites (1) rather than (2) as a reason for action is actually giving weight to sunk costs. For typically, in such circumstances it will be appropriate to interpret the argument on offer as an enthymeme, which, fully articulated, would resemble (2).²⁵

Indeed, if, as many have argued, our practice of interpretation is governed by principles of charity, won't it (almost?) always be reasonable to attribute a preference for redemption, or some similar preference, to an agent who is *apparently* giving weight to sunk costs? Here we encounter familiar worries about the trivialization of expected utility theory as a descriptive theory of human behavior. But the present case is not merely a special case of the general worry, I think, since there is often—as in the Vietnam example—some independent reason²⁶ for attributing the relevant preference.

One might, of course, argue that it is irrational to privilege a desire for redemption when doing so promises to result in losses that have not yet been incurred. Notice, however, that one who does so is once again engaged in the difficult task of arguing about how conflicts among competing values should be decided. And much of the allure—as well as the dialectical effectiveness—of identifying certain general patterns of arguments as fallacious lies in the promise that, by doing so, one can effectively circumvent difficult-to-arbitrate disputes about how competing values should be ranked.

There is, I believe, ample reason for skepticism about the idea that the rational requirement to ignore sunk costs is, when narrowly understood, a constraint with which much actual human behavior is inconsistent. We might have hoped otherwise. We might have hoped, for example, that a whole family of arguments for continuing costly endeavors might be dismissed out of hand, quickly and cleanly, as fallacious. If we are charitable, however, we should admit that few arguments seriously put forth are best understood as embodying this particular fallacy, and that we are unlikely to be spared the far messier and more difficult task of arguing about the value of competing ends, or about the worth of different goals.²⁷

Notes

¹ See, for example, Arkes and Blumer (1985), and the further references cited there, as well as Garland (1990). As will become clear below, the probative value of both the anecdotes and the studies is not beyond question.

² For representative examples, see Friedman (1986, pp. 279–280) and Nagl (1987, pp. 21–23).

³ In fact, there are two distinct forms of what is sometimes called “the sunk cost fallacy”. The first form manifests itself in a reluctance to abandon projects in which one has already invested considerable resources. This form—which has received more attention from psychologists and from those few philosophers to have addressed the topic of sunk costs—will be our exclusive concern in what follows. The second form is in some ways an opposite tendency,

inasmuch as it manifests itself in the *premature* abandonment of projects. Thus, many individuals will abandon a project once they realize that, given the past investments that have been made, the project has no hope of earning a net profit when taken as a whole. This conflicts with the advice that the economists offers, which is to ignore the past investments and continue the project if and only if it promises to be profitable from the present moment on. For good explanations of the distinction, see Steele (1996, pp. 608–610) and Friedman (1986, pp. 279–280).

⁴ Of course, the mere fact that someone appeals to sunk costs as a normative reason for pursuing a given policy does not show that he himself takes those sunk costs as constituting a genuine normative reason for pursuing that policy. If I believe that others tend to (mistakenly) treat sunk costs as reasons, then I might disingenuously appeal to sunk costs in attempting to influence their behavior.

⁵ The physicist Eugene Demler informs me that exactly parallel arguments were quite commonly made in the Soviet Union in the late 1980s in an attempt to justify continued Soviet involvement in Afghanistan.

⁶ It might be thought that questions about the rationality of continuing some course of action in order to avoid future regret cannot be so neatly disentangled from questions about the rationality of giving weight to sunk costs. For—it might be claimed—if one does not treat sunk costs as reasons for action, then one will have no reason to feel any regret at leaving the expensive ticket unused.

Concede, for the sake of argument, that one who does not treat sunk costs as reasons will have no reason to feel regret in the relevant circumstances. From this, of course, it does not follow that one will feel no regret after leaving the ticket unused—only that one will feel no regret *insofar as one is rational*. If I know that I will feel regret upon performing a certain action, then it is rational for me to take that regret into account in deciding whether to perform that action—even if such regret would be irrational (indeed, even if I *acknowledge* that such regret would be irrational). Compare, in a somewhat different context, Webber 1998, p. 114. In general, however, Webber is more skeptical than I am about the prospect of appealing to regret-avoidance in order to rationalize otherwise irrational behavior. (Webber does not specifically address the topic of sunk costs.)

⁷ As an anonymous referee pointed out to me, here (as elsewhere) the deficiencies in the psychological literature seem attributable to an overly behavioristic focus.

⁸ Nozick (1993, pp. 21–25). I borrow the terminology of “honoring” sunk costs from Nozick’s discussion.

⁹ Here and below, I will ignore the possibility that past commitments of this sort generate moral obligations which normatively constrain the agent’s future behavior. Although highly plausible in this particular example, the feature is incidental.

¹⁰ Of course, the best strategy for the pure expected utility maximizer might very well be to *mimic* the behavior of the pure honorer of sunk costs in an effort to win the advantages that accompany being perceived as one who honors sunk costs. Mimicking the behavior of the sunk cost honorer need not involve honoring sunk costs and indeed, might very well be enjoined by the maxim of maximizing expected utility. Whether such a strategy of deception is successful will depend not only on the skill of the mimicry but also on the perceptiveness of the opponent. I assume that the possibility of deception is not always available, and that (in some cases at least) there are limits to one’s ability to make one’s true policies opaque to others who have strong interests in correctly interpreting those policies. Difficulties in making one’s true policies opaque will be especially acute in democratic societies in which the content of the policies is at least somewhat responsive to the publicly manifested preferences of the citizenry.

¹¹ See Schelling (1960) and (1966). For a philosophical application of Schelling’s ideas, see Parfit (1984, Part I). Nozick notes the possibility of a connection between Schelling’s work and the topic of sunk costs in a passing remark (1993, p. 24).

¹² Moreover, there are possible situations in which the pure honorer of sunk costs fares better in systematic and predictable ways. In his critique of Nozick, Steele suggests that

Nozick's observations about the possible role that sunk cost honoring plays in offsetting weakness of the will are essentially uninteresting, inasmuch as "... any type of error might be compensated for by some other error to yield a desirable outcome. We can equally well imagine some circumstances in which supposing that $2 + 2 = 5$ would have a welcome outcome..." (Steele 1996: 608). No doubt, it is (almost always) a trivial matter to describe fortuitous circumstances in which an ostensibly irrational tendency or mistaken belief proves beneficial. What is not trivial, however, is to specify an interesting class of cases—the members of which are neither a hopelessly gerrymandered lot nor trivial variations of one another—in which an ostensibly irrational tendency proves advantageous in systematic ways, ways that might very well have been foreseen from the perspective of the agent herself.

¹³ Notice that in a society in which power is diffuse, the peculiarly bad policies in question need not be favored by any group or individual in the society (and indeed, might even be recognized *as* peculiarly bad policies by everyone) but arise naturally as the vector sum of competing views about what ought to be done. (Some cite sunk costs as reasons for staying the course until victory is achieved; others appeal to considerations of utility as reasons for getting out immediately; the divided nation ends up lingering for years and then pulls out without achieving any of its original goals.)

¹⁴ As Nozick speculates: "if someone offered us a pill that henceforth would make us people who never honored sunk costs, we might be ill-advised to take it" (1993, p. 23).

¹⁵ See especially Parfit's reply (2001) to Gauthier. The claim that a disposition might be irrational even if having that disposition best promotes an agent's substantive aims is perhaps most compelling when we consider dispositions to believe. Consider, for example, an athlete who benefits greatly, in predictable ways, from a disposition to form overly optimistic beliefs about his chances of success. If the athlete's degrees of confidence far outrun his evidence, then isn't it clear that he is being irrational, despite the fact that he benefits from the relevant disposition? However, even the claim that the athlete's overly optimistic beliefs are irrational is controversial. Richard Foley, for example, would deny that the athlete's beliefs need be irrational. (See his 1987 chapter 5.) Elsewhere, I have argued that the expected utility of holding a belief is irrelevant to the normative status of that belief (Kelly 2002).

¹⁶ On "the unity of a life", see also Alasdair MacIntyre's stimulating discussion (1984, ch.15). Elizabeth Anderson briefly connects the topics of sunk costs and "narrative unity" in her (1993, pp. 34–35).

¹⁷ In what follows, I will write as though the opportunity to satisfy some preference—or at least, the opportunity to satisfy some *rational* preference—provides one with a reason for action. I proceed in this way because such a view about practical rationality is typically presupposed by those who discuss sunk costs, especially those economists and psychologists who regard giving weight to sunk costs as a paradigm of irrational behavior. The argument of the present paper would, I believe, be no less (and no more) compelling if it were recast in a way that presupposes the truth of some alternative view. For recent development and defense of views according to which the notion of preference satisfaction does not play a central role, see Scanlon (1998) and Parfit (unpublished).

¹⁸ Here I simply gesture at (one) notion of a "redemptive action", without attempting to *define* that notion. Attempts at definition would have to accommodate the following kind of case. Suppose that I redeem some past efforts by performing act A_1 , but that, if I had not performed A_1 , I would have performed act A_2 instead, which also would have redeemed the same past efforts. Intuitively, act A_1 should count as a redemptive action, but it would be denied that status by condition (ii), since, if A_1 had not been performed, the efforts which it redeems would not have gone unredeemed but would have been redeemed by A_2 instead. (Here as elsewhere, analyses which make use of counterfactual locutions are vulnerable to the phenomenon of asymmetric overdetermination.) Of course, in the envisaged circumstances the fact that A_1 would be a redemptive action would not count as a reason for performing it instead of A_2 , since the latter action would also have this status.

¹⁹ Anderson endorses such “backward-looking” reasoning as legitimate and suggests that consequentialism is undermined by an inability to account for its status as such (1993, pp. 34–35). Sturgeon (1996) agrees that the reasoning is legitimate but argues that consequentialism can accommodate this fact (see especially pages 511–513). See also Anderson’s reply to Sturgeon in the same volume (pp. 541–542).

²⁰ A possible example: the preference for great pain on a future Tuesday to the mildest pain on any other future day. See Parfit (1984, pp. 120–125).

²¹ Compare “conservative” accounts of belief revision, according to which the mere fact that a proposition is presently believed makes it reasonable to continue to believe it in the absence of positive reasons for doing otherwise—even if it would not be reasonable to acquire the belief anew. For the development of such a view, see Harman (1986, ch.6). Bratman (1987) defends a somewhat analogous view with respect to intentions. Each of these views has, I think, more to be said in its favor than the one that I discuss in the text.

²² For emphasizing that there are important issues arising from the fact that the scope of the relevant actions can be demarcated in different ways, I am much indebted to the referee.

²³ Notice that if we employ the more inclusive terminology, then the norm of maximizing expected utility might actually require individuals with certain preferences to “honor sunk costs”. It would be better, I think, to avoid this.

²⁴ I include the parenthetical hedge in order to give due deference to the fact that the issue between Parfit and Gauthier discussed at the conclusion of Section 3 has not yet been decisively resolved. For if Gauthier is correct with respect to that particular issue, then the normative claim (2) might be open to challenge even if a narrow definition is employed.

²⁵ Because the conviction that people frequently do give weight to sunk costs in a way that is clearly irrational is widespread and well-entrenched, it is perhaps worth scrutinizing another putative paradigm of such behavior, in order to show that it too admits of a similar treatment. I have in mind the gambler who “throws good money after bad” in an effort to recoup his previous losses. (That is, in an effort to break even, the gambler continues to gamble, accepting bets that he would regard as unacceptable if not for his previous losses.) Here too, we have a case in which the continuing of a given activity is counterfactually dependent on past expenditures. Notice, however, that insofar as the gambler is driven by a desire to “break even”, he is *not* giving weight to sunk costs in the narrow sense. For the fact that winning these future bets would allow him to break even effectively changes the value of those bets from his perspective. That is, someone with a standing, independent preference for breaking even will be willing to accept certain bets when he is behind that he would be unwilling to accept if he were ahead, because the former (but not the latter) constitute a potential means to fulfilling this preference. Again, the behavior in question would only count as a case of honoring sunk costs in the narrow sense if neither this nor any similar preference were operative. In sum: although we often observe individuals “throwing good money after bad”, the hypothesis that these individuals are honoring sunk costs in the narrow sense is greatly underdetermined by the behavioral evidence that is typically available.

²⁶ That is, some reason which is independent of a desire to save expected utility theory as an adequate descriptive theory of human action.

²⁷ For helpful thoughts on earlier versions of this material, I am grateful to Derek Parfit, Robert Nozick, Noah Feldman, Luca Ferrero, and an anonymous referee for *Noûs*. For generous financial support, I am grateful to the Society of Fellows at Harvard University.

References

- Anderson, Elizabeth. (1993) *Value in Ethics and Economics*. (Cambridge, MA: Harvard University Press)
- Anderson, Elizabeth. (1996) “Reasons, Attitudes, and Values: Replies to Sturgeon and Piper”, *Ethics* 106: 538–554.

- Arkes, H.R. and Blumer, C. (1985) "The Psychology of Sunk Cost", *Organizational Behavior and Human Decision Processes* 35: 122–140.
- Bratman, Michael. (1987) *Intention, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press).
- Danto, Arthur. (1985) *Analytical Philosophy of History*. Reprinted under the title *Narration and Knowledge* (New York: Columbia University Press).
- Foley, Richard. (1987) *The Theory of Epistemic Rationality* (Cambridge, MA: Harvard University Press).
- Friedman, David. (1986) *Price Theory: An Intermediate Text* (Cincinnati: South-Western).
- Garland, Howard. (1990) "Throwing Good Money After Bad: the Effect of Sunk Costs on the Decision to Escalate Commitment to an Ongoing Project", *Journal of Applied Psychology* Vol. 75 (6) Dec: 728–731.
- Gauthier, David. (1997) "Rationality and the Rational Aim", in Jonathan Dancy (ed.) *Reading Parfit* (Oxford: Blackwell): 24–41.
- Harman, Gilbert. (1986) *Change In View* (Cambridge, MA: M.I.T. Press).
- Harman, Gilbert. (1999) "Practical Reasoning". Reprinted in his *Reasoning, Meaning, and Mind* (Oxford: Oxford University Press): 46–74.
- Kelly, Thomas. (2002) "The Rationality of Belief And Some Other Propositional Attitudes", *Philosophical Studies* 110: 163–196.
- MacIntyre, Alasdair. (1984) *After Virtue* (Notre Dame: University of Notre Dame Press).
- Nagl, Thomas. (1987) *The Strategy and Tactics of Pricing: A Guide to Profitable Decision-Making* (Englewood Cliffs, NJ: Prentice Hall).
- Nozick, Robert. (1993) *The Nature of Rationality* (Princeton: Princeton University Press).
- Parfit, Derek. (1984) *Reasons and Persons* (Oxford: Clarendon Press).
- Parfit, Derek. (2001) "Bombs and Coconuts, or Rational Irrationality" in Morris and Ripstein (eds.) *Practical Rationality and Preference: Essays for David Gauthier* (Cambridge, UK: Cambridge University Press): 81–97.
- Parfit, Derek. (unpublished manuscript) *Rediscovering Reasons*.
- Rawls, John. (1972) *A Theory of Justice* (Cambridge, MA: Harvard University Press).
- Scanlon, Thomas. (1998) *What We Owe to Each Other* (Cambridge, MA: Harvard University Press).
- Schelling, Thomas. (1960) *The Strategy of Conflict* (Cambridge, MA: Harvard University Press).
- Schelling, Thomas. (1966) *Arms and Influence* (New Haven: Yale University Press).
- Steele, David Ramsey. (1996) "Nozick on Sunk Costs", *Ethics* 106: 605–620.
- Sturgeon, Nicholas L. (1996) "Anderson on Reason and Value", *Ethics* 106: 509–524.
- Taylor, Charles. (1989) *Sources of the Self: The Making of Modern Identity* (Cambridge: Harvard University Press).
- Velleman, David. (2000) "Well-Being and Time" reprinted in his collection, *The Possibility of Practical Reason* (Oxford: Oxford University Press): 56–84.
- Weber, Michael. (1998) "The Resilience of the Allais Paradox", *Ethics* 109: 94–118.

Copyright of *Nous* is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.