

The Sunk Cost “Fallacy” is Not a Fallacy.

Ryan Doody

November 1, 2013

Abstract

Business and Economic textbooks warn against committing the *Sunk Cost Fallacy*: you, rationally, shouldn't let unrecoverable costs influence your current decisions. In this paper, I argue that this isn't, in general, correct. Sometimes it's perfectly reasonable to wish to carry on with a project because of the resources you've already sunk into it. The reason? Given that we're social creatures, it's not at all unreasonable to care about wanting to act in such a way so that a plausible story can be told about you according to which your diachronic behavior doesn't reveal that you've suffered, what I will call, *diachronic misfortune*. Acting so as to hide that you've suffered diachronic misfortune involves striving to make yourself easily understood to others (as well as your future self) while disguising any shortcomings that might damage your reputation as a desirable teammate. And making yourself easily understood to others while hiding your flaws will, sometimes, put pressure on you to honor sunk costs.

1 Introduction

Conventional wisdom, as well-documented in introductory Business and Economics textbooks, holds that it's irrational to commit the *sunk cost fallacy*. Very roughly: you commit the sunk cost fallacy when you let unrecoverable costs influence your current decision-making.¹

Economists and Business Majors notwithstanding, most of us *do* commit the sunk cost fallacy.² (For the sake of picking a more neutral phrase, let's follow KELLY 2004 and refer to this behavior as *honoring sunk costs*.) Sunk Cost cases range from the mundane to the profound, from the personal to the political. Here's one example: You bought a non-refundable, non-transferable opera ticket — but, by the time the night of the show rolls

¹This is intentionally rough. We will work to spell this out more precisely in a moment.

²For a collection of psychological studies to this effect, see ARKS AND BLUMER 1985. For a collection of anecdotal evidence, please consult my mother.

Also, Econ and Business students appear to honor sunk costs with the same gusto as the rest of us. Learning about the fallacy seems to have little affect on one's propensity to commit it. See: ARKS AND BLUMER 1985, GARLAND 1990.

around, you are no longer sure you want to go. Here's another less-mundane example: you've devoted many years of your life to a career in Finance — but, after years spent advancing up the corporate ladder, you are no longer sure that this is a job you enjoy doing. And here's another, this time more-political, example: We expend considerable resources (as well as sustain significant casualties) fighting a war — which now seems to many to be almost unwinnable. There are, of course, a myriad of other examples. In each of these situations, it's hard not to think things like, e.g., “but I've already spent money on this,” or “but all that time and work will have been for nothing,” or “if we don't keep fighting, those who've fallen during combat will have died in vain!”

There are lots of cases in which we feel pressure to honor sunk costs. But it is not true that *whenever* we've sunk some costs into a endeavor we feel pressure to carry on with it.³ Here's an example: You buy fire insurance for your house and your house doesn't burn down. But there is no pressure whatsoever to honor the costs you've sunk into the insurance premiums by, for example, burning your house down. So *sometimes* we feel the “pull” to honor sunk costs, but sometimes we don't. What's the difference? And, in those cases in which we *are* tempted to honor sunk costs, what's so irrational about succumbing to it? In order to make a case, one way or other, about the rationality of honoring sunk costs, we need to get clearer about exactly why we feel the pressure to do so when we do.

In this paper, I am going to do two things. First, I am going to provide an account of what it is that makes the difference between those cases in which we feel pressure to honor sunk costs and those cases in which we don't. Second, *contra* conventional wisdom, I will suggest that once we come to understand *why* we feel the pressure to honor sunk costs, it's no longer clear that doing so is irrational. Here's the idea: In the cases in which we feel pulled to carry on with a project because of the costs we've sunk into it, the honoring of sunk costs allows us to *maintain plausible deniability about having suffered, what I call, diachronic misfortune*. And the desire to maintain plausible deniability about having suffered a diachronic misfortune — wanting to be able to spin a plausible autobiographical tale that casts its protagonist in a flattering light — is a nearly-universally-had and deeply-rooted one. It is a desire that proverbially resides close to our proverbial hearts; it's central to who we are. In fact, given the kind of creatures we are — social, deeply reliant our ability to effectively coordinate — it's not at all unreasonable to expect creatures like us, via a process of social evolution, to come to internalize the desire to tell exonerating stories about ourselves.

Here's how I will proceed. In the next sections, we will get clearer both about what it *is* to honor sunk costs, and why we feel the pressure to do so in some cases but not others. I will defend **Claim I** — we feel tempted to honor

³This isn't just to say that we feel pressure to do so which is ultimately *outweighed* by other considerations — there is *no* temptation to honor sunk costs whatsoever.

sunk costs when there's an asymmetry in the prospects of telling yourself a plausible diachronically-flattering story between carrying on with the Sunk Cost Project and abandoning it. Next, I will suggest that it isn't always irrational to honor sunk costs by arguing for **Claim II** — it's reasonable to expect social creatures to care, profoundly, about this type of self-serving autobiographical storytelling, because to do so promotes our social fitness.⁴

2 What is it to Honor Sunk Costs?

So far, I've given only a very rough characterization of what it is to honor sunk costs. Allow me to rectify that using an example.

A Night at the Opera? It's Saturday night and you have a ticket to *La Traviata*. You bought the ticket in advance, two weeks ago. (Let's say, for the sake of the story, you paid \$100). Thing is: *you can't decide whether or not to go*.

Two weeks ago — when you were buying the ticket — you wanted to go. But now you're not so sure. “The opera,” you think “would be nice — but staying home would be nicer.” In fact, the following is true of you:

were you to have, say, found this ticket — rather than spent your hard-earned money on it — it'd be a no-brainer: you'd stay home.

But, alas, things aren't that simple. “Look,” you think, “I could have just as easily *not* bought that ticket, saved myself the money, and stayed home with \$100 in my pocket.” If only! You can't undo what's been done. Your available options are clear: either *go* or *stay*. What to do?

Let me make the story a bit clearer by representing it with a couple tree-diagrams:

⁴Is this a bait & switch?: I draw you in with the promise of rationalize honoring sunk costs, but really end up rationalizing something else instead. (You don't, for example, successfully rationalize *poking yourself in the eye* by arguing that in some cases — ones, for example, in which someone offers you a very very large sum of money if you poke yourself in the eye — it's rational to do so).

It's not. Here's why. The paradigmatic cases of sunk cost honoring are constitutively linked, in a strong way, to maintaining plausible deniability about your diachronic failings. When they are linked, the connection is strong — it's not just a superficial, fragilely contingent causal link. Furthermore, you can understand the thesis this way: many of the cases, in fact the paradigmatic ones, are such that the behavior therein is *not* irrational. If you think, though, that (part of what my argument shows is that) these are *truly* cases of honoring sunk costs, then — fine. Understand me, then, as saying that we, as a matter of fact, rarely ever commit the sunk cost fallacy. If, on the other hand, you *do* think the commonly cited examples of sunk cost honoring are cases in which we truly do honor sunk costs, then the point I will argue for stands: it is not irrational to care about sunk costs.

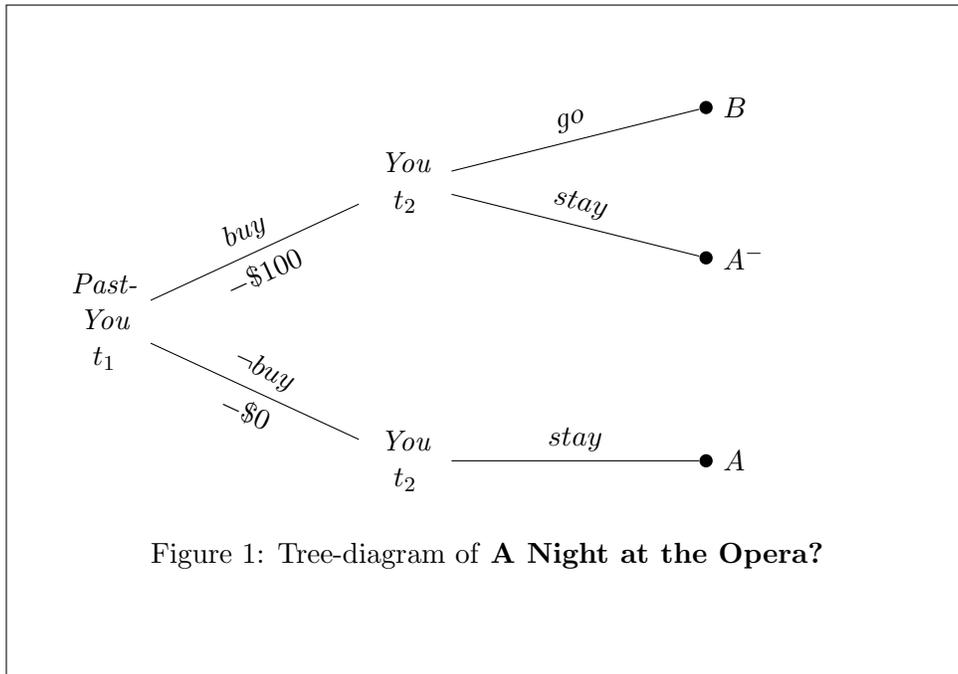
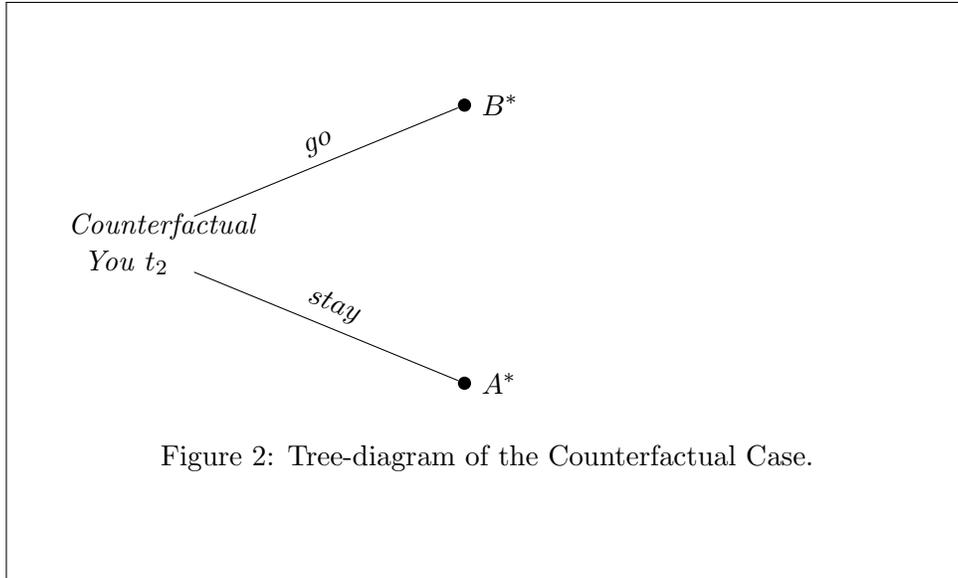


Figure 1: Tree-diagram of **A Night at the Opera?**

Reading Figure 1. At t_1 , you faced the decision of whether to spend \$100 on the opera ticket (go up) or to not buy the opera ticket (go down). If you buy the opera ticket, then, at t_2 , you will face the decision of whether to go to the opera (go up), thus bringing about outcome B , or to stay home (go down), thus bringing about outcome A^- . If you don't buy the opera ticket, then, at t_2 , you will have no choice but to stay home. This is outcome A : an outcome that's *almost just like* outcome A^- were it not that you were \$100 richer.



Reading Figure 2. Were you to have, say, found the opera ticket — rather than spent your \$100 on it — then, at t_2 , you'd face the decision of whether to go to the opera (go up), thus bringing about outcome B^* , or to stay home, thus bringing about outcome A^* .

In cases like this one, I feel pressure to go. Had I *not* bought the ticket — had I stumbled across it, or were it to be Free Opera Night, or something like that — and I didn't feel like going, I wouldn't go. Having a pattern of attitudes like this is characteristic of honoring sunk costs.

Sunk Costs: You *honor sunk costs* if you decide to ϕ rather than ψ , but are disposed, had sunk costs not been sunk, to ψ rather than ϕ .

You, like me, might feel tempted to honor sunk costs in **A Night at the Opera?** — you might feel pressure to *go* rather than *stay*, even though you're disposed, had you you not sunk \$100 into the project of going to the opera, to *stay* rather than *go*. But why? What's the difference between the two cases? Here's an obvious suggestion: you feel pressure to carry on with a project when unrecoverable resources have been lost to the project. If you've *bought* the opera ticket, then you've sunk some unrecoverable resources into the project of going to the opera. On the other hand, if you *found* the opera ticket (by stumbling across it), then no resources have yet been expended on the opera-going project. This suggestion is not quite right, however, as the following example brings out.

Short-Changed at the Opera. You have little to no desire to go see *La Traviata* two weeks from now. (You have no strong definitive desires about what to do in two weeks at all). And you,

certainly, have no intention to buy a \$100 opera ticket. In fact, your trip to the Opera Company’s ticketing booth had nothing to do with the opera at all — you had a very rare \$1000 bill in your pocket that was desperately in need of breaking.

Correctly assuming that the Opera Company would be able to break your bill, you approached a Ticket Booth Agent. Unbeknown to you — and, much to your misfortune, unnoticed by the absent-minded Ticket Booth Agent — the (absolutely nonrefundable-under-any-circumstances) tickets for next fortnight’s production of *La Traviata* eerily resemble \$100 bills. You realize much too late that the Ticket Booth Agent mistakenly gave you nine \$100 bills and one ticket to the opera. What luck! Ugh!

Fast-forward two weeks. It’s Saturday night. You don’t really feel like going to the opera tonight. You’d rather stay in and enjoy a relaxing evening in front of the TV. You think to yourself, “it’s a shame that I got shorted \$100 by that Ticket Booth Agent, but there’s nothing (short of issuing a formal complaint with his superiors) that I can do about it now.” What to do?

In the first case (**A Night at the Opera?**), I would feel considerable pressure to go to the opera. In the second case (**Short-Changed at the Opera.**), I wouldn’t.⁵ But in *both* cases, an unrecoverable \$100 has gone toward the opera-going project. So the pressure we feel isn’t owing *just* to the loss of money. The important difference between the two cases is that in the former, but not the latter, the money was sunk into the opera-going project *intentionally* — the opera ticket was acquired *on purpose* in **A Night at the Opera?** and *accidentally* in **Short-Changed at the Opera.**

Here’s why this difference is important. In both cases, there has been an exchange of money for a ticket. But *buying* the ticket, as opposed to acquiring it by accident, reveals a preference, at time t_1 , for *buying* over *not buying*.⁶ This means that: at time t_1 , your beliefs and desires were such that the *expected utility* of purchasing the opera ticket exceeded the *expected utility* of not purchasing it.⁷ The outcome that will result from your decision at

⁵Or at least I would feel considerably *less* pressure to go in the latter than in the former.

⁶This assumes that you were behaving rationally at time t_1 . If we turn off this assumption — or we amend the case so that, at time t_2 , it is reasonable for you to believe that you, at time t_1 , were not behaving rationally — we get a more complicated case. It’s best, at this point, to simplify by making this assumption.

⁷Choosing to buy the ticket is — in all relevant respects — to place a bet. Namely, a bet that pays out a night at the opera, if the world turns out one way, and pays out a night spent at home, if the world turns out a different way. The vast *vast* majority of our actions can be understood as a form of betting. Insofar as the outcome of an action turns on how the world might turn out to be, taking that action is to take a bet. This idea is the cornerstone of Bayesian Decision Theory.

time t_1 turns on what will happen — what you will choose to do, and what the world will be like — at time t_2 . Acquiring the ticket intentionally reveals information about how you-at-time- t_1 believed and wanted the world to be. Acquiring the ticket accidentally reveals nothing about what your beliefs and preferences were like at time t_1 .

Exactly what your decision to buy the ticket reveals about your beliefs and preferences-over-outcomes very much depends on how you-at-time- t_1 conceived of it. This can be illustrated by telling two different versions of the story, like so:

A Night at the Opera? (Binding). You long to be someone who regularly goes to the opera. You aspire to be the kind of person who appreciates High Culture. As it is, though, you aren't that kind of person at all. You find the opera (as well as: the ballet, modern art museums, French films, free verse poetry, etc., etc.) to be tedious and boring. Consequently, you know that — left to your own devices — you will never go to the opera, you will never develop a taste for the finer things, and you will eventually die without ever coming to appreciate High Culture. You don't want that to happen, though.

It is in *that* spirit that you approach the Opera House's ticket booth. You purchase a ticket for *La Traviata* for two weeks from now *because you want your future-self to go to the opera*. You think: "What I *really* want is to *want* to go to the opera. And, given that I probably won't come to want to go to the opera out-of-the-blue, the best way to get myself to want to go to the opera is to make myself go." So, at time t_1 , you prefer that future-you goes to the opera whether future-you feels like going or not.

A Night at the Opera? (Betting). You decide to purchase a ticket for *La Traviata* — not because you want future-you to go to the opera *come what may* — but instead to give yourself the *option* to go to the opera if you decide you want to go.⁸

In *both* versions of **A Night at the Opera?** there's pressure to honor sunk costs by opting to *go* rather than *stay*. But what the purchasing of the opera ticket at time t_1 *says* about your beliefs and preferences at that time depends on the version of the story. In **Binding** (which is the version that

⁸The decision to purchase the ticket is like taking a bet that turns on whether or not you will feel like going. It is not an essential feature of the case, however, that this "bet" turns on *how you will feel about going* rather than, say, the weather. For example, you might be the opera ticket with the intention of going *unless there's heavy snowfall that evening*. It's a pain to go out when it's really coming down out there. And yet even if it does snow Saturday evening, there's still pressure to honor sunk costs by going.

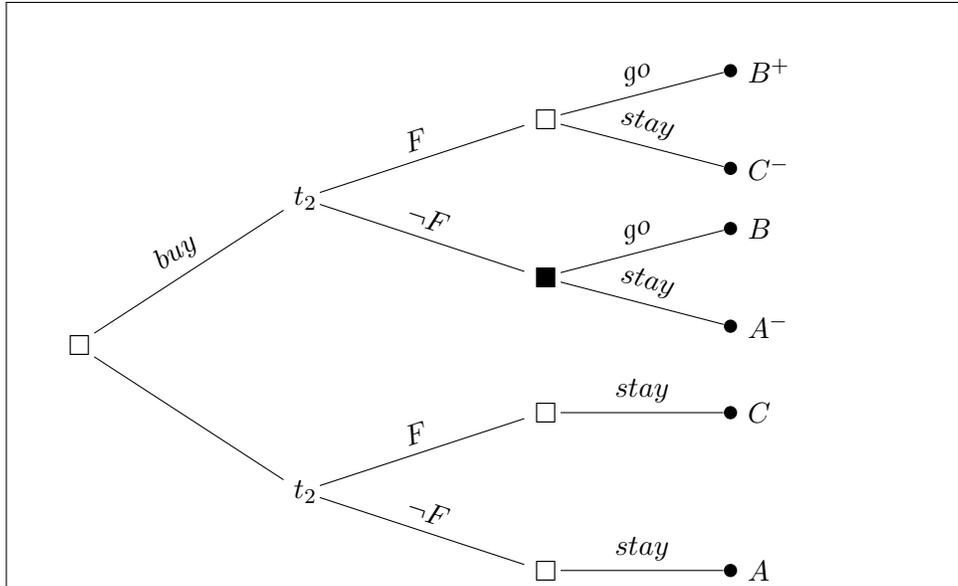


Figure 3: Tree-diagram of **A Night at the Opera? (Betting)**.

Reading Figure 3. You buy the ticket in order to give your future-self *the option* to go. At time t_1 , you have to decide whether or not to *buy* the opera ticket. At time t_2 , you will learn whether or not you *feel like going*. (Where F stands for *feels like going* and $\neg F$ stands for *doesn't feel like going*). After learning whether or not you feel like going, you-at-time- t_2 must decide whether to *go* to the opera or *stay* home. Of course, if you decided at time t_1 to *not buy* the ticket, you have no choice but to *stay* at time t_2 .

is implicitly evoked in the tree-diagram of Figure 1), you *unconditionally* desire that future-you goes to the opera.⁹ Your buying of the ticket, in this case, is being used as a way to *bind* your future-self. We do sometimes have preferences like this. Consider, for example, buying a year-long gym membership. Often, when people purchase gym memberships they don't just want to give their future-selves *the option* to go work out if they so choose — rather, they want their future-selves to *go work out* whether they feel like it at the time or not.

⁹This isn't exactly right. It's rare that we prefer one thing to another *come what may*. Even here, you presumably don't desire future-you goes to the opera *no matter what*. For example, if the Apocalypse begins Saturday night, you probably desire that future-you do something more exciting than spend the night at the opera. There are countless other conditions your opera-going desire turns on. The sense, then, in which your desire to go to the opera in **Betting**, your desire in **Binding** is unconditional. The difference is that your preferences in **Binding** are unconditional with respect to how you'll feel in the future, whereas your preferences in **Betting** are sensitive to how future-you will feel.

Sometimes our preferences are like those described in **Binding**, but at least just as often our preferences are more like those described in **Betting**: we want our future-selves to do what they feel like doing. Purchasing the ticket, in this case, gives your future-self *the option* to go to the opera (if she or he so chooses). What does *intentionally* exchanging \$100 for an opera ticket reveal about your beliefs and preferences?

1. **Your Beliefs and Preferences in Binding.** At time t_1 , you preferred buying the opera ticket over not buying it. Because the outcome in which you stay home *and* are out \$100 (outcome A^-) is clearly worse, by your own lights, than the outcome in which you stay home but never bought the ticket in the first place (outcome A), you-at-time- t_1 must have most preferred the outcome in which you spend \$100 on a ticket *and go to the opera* (outcome B). Otherwise, your desire to purchase the opera ticket at time t_1 wouldn't have been rational.¹⁰ So you-at-time- t_1 must prefer outcome B to outcome A to outcome A^- .
2. **Your Beliefs and Preferences in Betting.** When your preferences are conditional on *how you will feel*, your decision at time t_1 to purchase the opera ticket reveals less about your preferences-over-outcomes than it does in **Binding**. It is consistent with the rational purchase of the opera ticket at time t_1 that you-at-time- t_1 *do not* prefer outcome B (the outcome in which you've bought the opera ticket and go, despite not feeling like it) to outcome A (the outcome in which you haven't bought the ticket, don't feel like going, and don't go). However, purchasing the ticket *does* reveal something about your beliefs at time t_1 , given some commonsense constraints on what it's plausible to care about.¹¹ In particular, in order for your decision at time t_1 to be rational, you must think that it *reasonably likely* that at time t_2 you will feel like going. To illustrate, notice that if at time t_1 you were *extremely* confident in $\neg F$ — that at time t_2 you won't feel like going to the opera — then (because you prefer A to both A^- and B) the

¹⁰**Proof.** Your decision to buy the ticket rather than not is rational only if the *expected utility*, for you, of buying the ticket is greater than the *expected utility* of not buying. We can think of *buying a ticket* as placing a bet — a bet that, in this case, turns on what you will decide to do at time t_2 . The bet pays off a night at the opera, minus the cost of the ticket, if you end up deciding to go. And it pays off a night at home, minus the cost of the ticket, if you end up deciding to stay. Not taking the bet, however, results in a sure-thing night at home. Purchasing the ticket, then, is rational only if the weighted average of the value, for you, of outcome B and the value, for you, of outcome A^- (where the weights correspond to how confident you are that you will go) is greater than the value of outcome A . Because outcome A^- is, by your lights, worse than outcome A , you must prefer outcome B to outcome A in order for the purchasing of the ticket to be rational.

¹¹The constraints are as follows: (1) You, *ceteris paribus*, prefer to do what you feel like doing; (2) You, *ceteris paribus*, prefer remaining \$100 richer; and (3) You, *ceteris paribus*, prefer to go to the opera if you feel like going over staying home if you feel like staying home. (Despite (1), we should hold off on ranking A^- ahead of B because I will argue that *all else is not equal* in this case).

act of purchasing the opera ticket would be akin to simply throwing money away! So you have to have at least some positive credence that you will feel like going. And, in fact, you have to be a great deal more confident that that.

When Saturday night rolls around, you have two available options: you can decide to *stay home* or *go to the opera*. As much as you might wish otherwise, there is no option available to you that would, were you to take it, result in outcome A — you cannot now go back in time and prevent you-at-time- t_1 from purchasing the opera ticket. Outcome A is no longer accessible to you. But, as it turns out, it *was* accessible to you. Let me introduce some terminology:

An outcome O is *diachronically accessible to you*, at a time t_i , just in case you faced a choice, or series of choices, prior to time t_i such that were you to have chosen differently at that time, outcome O would have resulted.

Saturday evening, outcome A is diachronically accessible to you. By opting to *stay home* you will bring about outcome A^- which is clearly worse, by your own lights, than outcome A . If you opt to *stay home*, you suffer, what I will call, *diachronic misfortune*.¹²

Misfortune You've suffered *diachronic misfortune* if you've made a series of decisions that resulted in an outcome O such that there is another outcome O' that (1) is diachronically accessible to you, and (2) is better, by your own lights, than O .¹³

Perhaps then, we feel pressure to honor sunk costs when *not* doing so would result in the suffering of diachronic misfortune. In both versions of **A Night at the Opera?**, if you decide to not *go to the opera*, you will suffer diachronic misfortune. But in **Short-Changed at the Opera**, if you decide to not go, you won't thereby suffer a misfortune of this sort.

¹²**Important Warning:** It takes *very very little* to suffer diachronic misfortune. One can act perfectly rationally — one can do absolutely everything one, rationally, should do at each time — and still, as a result of bad luck, end up in a sub-optimal outcome relative to some other outcome that's diachronically accessible to you. Suffering diachronic misfortune is totally consistent with being impeccably rational. (Perhaps, there are *some* kinds of diachronic mistakes — like, for example, having so-fickle-as-to-be-money-pumpable preferences — that *are* irrational. For the purposes of this paper, however, I wish to remain completely neutral about this).

¹³This 'better, by your own lights' business is a bit tricky. Because we're dealing with a series of decisions — which are extended in time — it is possible that your preferences change as time goes by. If so, it's not clear that there's a temporally-absolute fact of the matter about which outcomes are better, by your own lights, than others. In all the cases I'm interested in, there will be an outcome that is *at all times* worse, by your own lights, than some other outcome. But it might not be the case that it is at all times worse than *the same* outcome.

This suggestion cannot be right, either. We sometimes *do not* feel any pressure to honor sunk costs even when not doing so would result in the suffering of a diachronic misfortune. Here is an example.

Camping Rainstorm.¹⁴ You were planning a camping trip. The weather forecast had it that it was likely to rain. Reasonably, then, you decide to rent some rain-gear — including a fairly expensive raincoat. You bring your new rain-gear, as well as all the other camping necessities, along with you on your trip. The weather forecast, however, turns out to be incorrect: there’s not a cloud in the sky. Nevertheless, you *could* still don the fairly expensive raincoat. After all, you spent all that hard-earned money on it! Wearing the raincoat, of course, won’t keep you any drier (you’ll be water-free no matter what you wear) and you’re sure you’ll feel pretty silly walking around with a completely ineffectual raincoat on. What to do?

The decision to *wear* the raincoat in **Camping Rainstorm** seems totally nuts. There is absolutely no pressure to do so. What’s the difference between (i) buying a ticket, learning that you don’t feel like going, and going to the opera anyway; and, in the second case, (ii) buying a raincoat, learning that there will not be a rainstorm, and wearing the raincoat anyway? The desire to avoid suffering diachronic misfortune cannot, at least, be the whole story.

3 Honoring Sunk Costs and The Spinning of Your Social Story

There are cases in which we hear the siren’s call of our past expenditures, lulling us toward one course of action over another. There are other cases, too; cases in which the call of our sunk costs falls on deaf ears: we feel little to no pressure to honor sunk costs.¹⁵ Why do we feel pressure to honor

¹⁴This case is partially inspired by an example of Caleb Pickard’s, from his comments on an earlier version of this paper delivered at the Rocky Mountain Philosophy Conference. Caleb’s example involves the decision to buy a *Put Option* — a financial contract that allows the buyer to lock-in the current price of some good. If, for example, you believe that the price of, say, corn is likely to rise, you could buy a *Put Option on Corn* that would allow you, in the future, to purchase corn at its current price. Suppose that you buy the Put Option. But, unforeseen events transpire, and the price of corn actually plummets. It’s now time t_2 : you can opt to *use your Put Option* — and buy a quantity of corn at its earlier, higher price — or opt to *buy corn without the Put Option* — and buy the quantity of corn at its current, lower price. Caleb correctly pointed out that there is no temptation to honor sunk costs (by using the Put Option) in this case — and that to do so has the strong stench of irrationality.

¹⁵There are, too, cases in between. Cases in which, to stretch the already-somewhat-tired metaphor a bit more, the siren’s call of our sunk costs can be heard, but are decisively drowned-out by ambient noise — in other words: cases in which we have some reason to

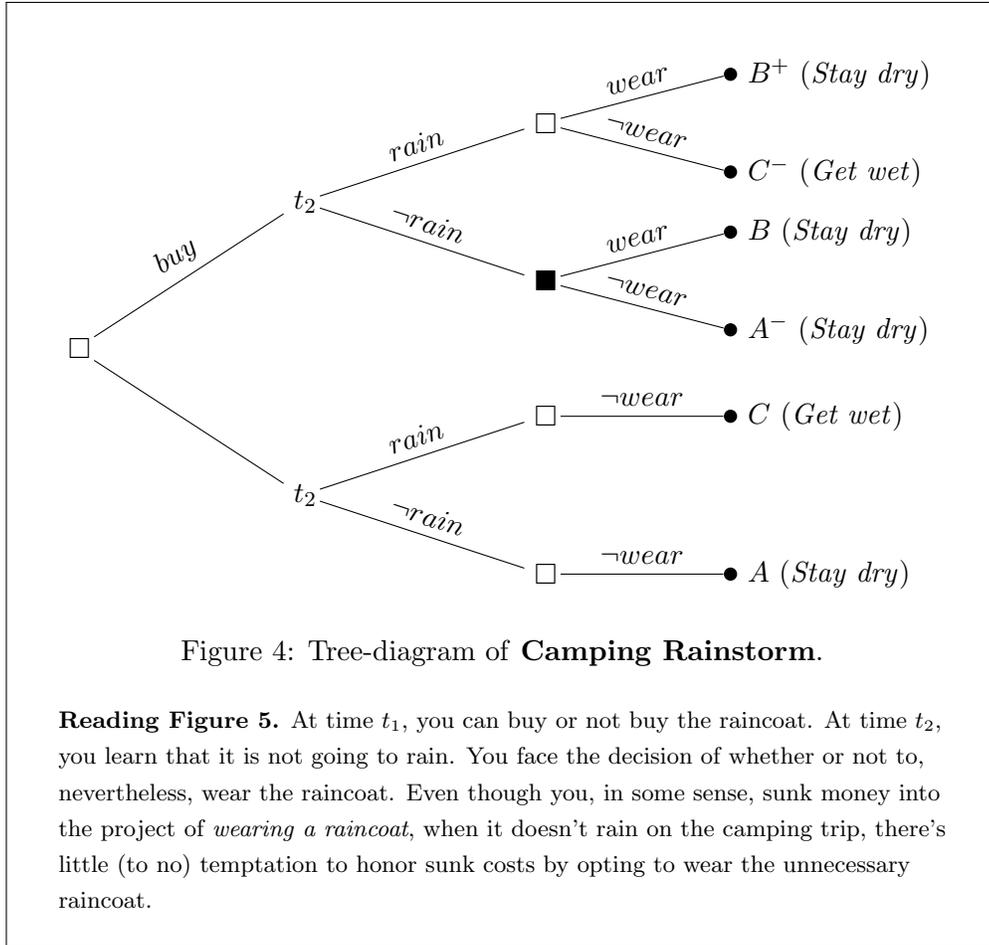


Figure 4: Tree-diagram of **Camping Rainstorm**.

Reading Figure 5. At time t_1 , you can buy or not buy the raincoat. At time t_2 , you learn that it is not going to rain. You face the decision of whether or not to, nevertheless, wear the raincoat. Even though you, in some sense, sunk money into the project of *wearing a raincoat*, when it doesn't rain on the camping trip, there's little (to no) temptation to honor sunk costs by opting to wear the unnecessary raincoat.

sunk costs in some cases but not others? Here's my hypothesis. The cases in which we such pressure are ones in which it will be easier to integrate the action that *honors sunk costs* into a plausible autobiography according to which its protagonist has not suffered diachronic misfortune. In these cases, there will be an asymmetry in the prospects of spinning a plausible story that casts you in a good light; in the cases in which we don't feel pressure to honor our sunk costs, however, honoring sunk costs will make the prospects of telling an exonerating story *just as dire* as they would be were you to not honor sunk costs.

honor sunk costs, but that reason is entirely swamped by other considerations. Imagine, for example, a case much like **A Night at the Opera?** except that, come Saturday night, you become ill. You don't feel like going to the opera because you are sick — the thought of being anywhere but in bed, an arms-length away from a box of Kleenex seems downright dreadful! This is a case in which, although you might feel some sunk-cost-related pressure to *go*, you would find being at the opera while ill **so** unpleasant that it's overwhelmingly clear to you to *stay home* — preferably, in bed, with a cup of soothing tea.

Claim I: You will feel pressure to *honor sunk costs* when:

- (1) There's no plausible story to be told about your behavior according to which you
 - (a) sink some costs into a project,
 - (b) later, abandon that project, and
 - (c) haven't suffered diachronic misfortune.

But,

- (2) If you carry on with the project, it *is* possible to tell a plausible story according to which you haven't suffered diachronic misfortune.

Here's the claim: You don't want to act in such a way that it will be obvious that, either, (a) you have made a bet and *lost* or (b) your preferences are diachronically unstable. Weakness is unbecoming. If it is obvious, either, that you lost a bet or that you have fickle preferences, you project weakness.¹⁶ We feel compelled to *honor sunk costs* when doing so will aid in hiding that we've suffered a diachronic misfortune. Of course, sometimes our shortcomings will be impossible to hide. In *these* cases honoring sunk costs loses its appeal.

We all make mistakes — diachronic and otherwise. We try not to, but mistakes happen (despite, and sometimes because of, our best efforts to avoid them). If you'd like to avoid making a mistake, but can't, you might still be interested in at the very least *covering up* that you've made a mistake. What does this have to do with sunk costs? The claim is that it is *this* want — the desire to conceal, as much as possible, our mistake-making — that, at least in some cases, motivates us to honor sunk costs.

3.1 A Night at the Opera? Binding and Betting

In both versions of this story, opting to *stay* reveals that you've suffered diachronic misfortune. Given that you've already bought the ticket, were you to *stay*, you'd bring about outcome A^- which is worse — clearly, and undeniably — than outcome A . And, at time t_2 , outcome A is diachronically accessible to you. Therefore, were you to *stay* rather than *go*, there would be no plausible story that could be told about your behavior according to which you haven't suffered diachronic misfortune.

Furthermore — and just as crucially — in both versions, if you opt to *go*,

¹⁶All weakness is undesirable, but some kinds might be worse than others (e.g., the diachronic instability of your preferences, lacking *self*-awareness).

a plausible story *can* be told about you according to which you remain misfortune-free. Here’s why. In **Binding**, if you opt to *go*, you can successfully hide that you’ve had a change of heart. Your preferences *have* changed — and there’s nothing you can do about that now. (You-at-time- t_1 preferred B to A but, now, prefer A to B). But, because your preferences with respect to outcomes B and A are *inert* (you are no longer in a position to bring outcome A about), it is possible for you to disguise the preference change by going ahead with the Opera Project you, at time t_1 , sunk costs in to. Similarly, in **Betting**, if you opt to *go*, you can successfully hide that you’ve lost a bet about how you would feel. By bringing about outcome B you make a diachronic mistake — B is worse (and clearly so) than A . It’s generally accepted that it is *ceteris paribus* worse to do something you don’t feel like doing. But, because *how you feel* is non-public, you are able to hide the fact that you don’t feel like going by opting to *go*.

In short, the following is true in both versions of the story.

- If you decide to *stay*, then (given that you bought the ticket) there’s no plausible story to tell according to which you haven’t made a diachronic mistake.
- If you decide to *go*, however, there *is* a plausible story to tell according to which you haven’t made a diachronic mistake.

If you decide to *go*, your behavior — first, buying an opera ticket, then going to the opera — is consistent with *a story* in which you are diachronically virtuous. It’s true that your action now cannot make it any less true that your preferences have changed, or your prediction didn’t pan out — but, by deciding to *go*, you can effectively *hide* these mistakes.¹⁷

On the other hand, In **Short-Changed at the Opera**, there is nothing about your acquisition of the opera ticket that would make it reasonable for anyone to infer anything substantive about, either, what your preference over the relevant outcomes were, or how likely you took it to be that you would feel like going to the opera Saturday night. To wit, it’s completely compatible with you *accidentally* acquiring the ticket that you all-along preferred A^* to B^* and were maximally confident that you wouldn’t feel like going to the opera on Saturday. Therefore, your decision at time t_2 will not reveal anything at all about the diachronic stability of your preferences (or lack thereof), or bad bet making — no matter what you decide, there will be a coherent (and plausible) story to tell according to which your preferences have the virtuous feature of being diachronically stable and your bets have all turned out favorably.

¹⁷The claim isn’t that by deciding to *go* you will redeem yourself by somehow undoing your diachronic mistakes; rather, the claim is that by deciding to *go* you can attempt to *hide* your failings. It’s *this* asymmetry — the asymmetry in the prospects for telling a plausible social story according to which you haven’t made any diachronic mistakes — that gives outcome B a leg up over outcome A^- .

3.2 Camping Rainstorm

In *this* story, however, no matter what you do at time t_2 — opt to wear the raincoat or opt not to — you will not be able to maintain plausible deniability about having suffered diachronic misfortune. The reason is this:

You’ve rented a raincoat and it didn’t rain, so you’ve lost a bet. If you decide to *not wear* the raincoat, there’s *no* plausible story that can be told about your behavior in which you haven’t messed up somehow. Why? Because the outcome (which we’ve been calling A^-) in which you rent the raincoat, it doesn’t rain, and you don’t wear it is sub-optimal — it’s worse than the outcome A in which you *didn’t* rent the raincoat, it doesn’t rain, and so you don’t wear it.¹⁸ And, more importantly, if you decide to *wear* the raincoat anyway — despite the fact there’s no rain — there’s also no plausible story that can be told about you according to which you haven’t messed up somehow. Why? Because, first, it is obvious that it isn’t raining. The weather is very public. So, there is no plausible story about your behavior in which it rains. And, second, people don’t wear raincoats when it’s not raining. So, it’s natural to suppose that when you purchased the raincoat at time t_1 , you had had conditional preferences — you didn’t want Future You to wear the raincoat come what may. And so, were you to wear the raincoat, you’d *still* be signaling that you’d lost a bet.¹⁹ Given that it didn’t end up raining, it would have been better for Past You to have not bought the raincoat.

3.3 Plausible Deniability

In order for you to maintain plausible deniability about something, you have to construct a narrative about your behavior that’s *plausible*. But what is it for a narrative to be plausible? And to whom are we constructing our narratives for?

You will be *not* be able to construct a plausible narrative about your behavior according to which you haven’t suffered diachronic misfortune when it is *obvious* that you’ve taken an action that has resulted in an outcome O which is sub-optimal relative to an outcome that’s diachronically accessible to you. If you want to tell a plausible story, there are two ways to do it. First, if it is *obvious* that O is sub-optimal, you might yet be able to maintain plausible deniability by *misrepresenting* O as some other outcome. This can be accomplished if the state-of-the-world that partially constitutes

¹⁸Think about Counterfactual You, hanging out in the possible world in which you decided against renting the raincoat, who’s enjoying the beautiful weather, raincoatless (just like Actual You) but who is, also, the-cost-of-a-fairly-expensive-raincoat richer than you. There’s no way to spin your behavior into a story according to which you inhabit the best-of-all-possible branches of the decision-tree.

¹⁹If anything, by wearing the raincoat when it isn’t raining, you are *loudly broadcasting* that you lost a bet. It’s as if you are yelling: “I BOUGHT A RAINCOAT, SEE? AND, LOOK, IT DIDN’T RAIN! LOOK AT ME! I MESSED UP! WHOOPS!”

O is suitably *non-public*. Second, if it is *obvious* that outcome O is the outcome your actions have brought about, you might yet be able to maintain plausible deniability by *disguising* the fact that you prefer a diachronically accessible outcome to O .

What makes a story about your behavior *plausible*? In order for the narrative to be plausible, it's not enough that your diachronic behavior merely meets some formal constraints. The story must also attribute attitudes to you that seem reasonable. What counts as "plausible" will depend on the kinds of things that we around here consider to be relatively natural to care about.

For whom are constructing these narratives? Our stories are partially directed toward the other members of community, and partially directed toward ourselves. As a heuristic (because it is not always possible to tell who's watching when), we might find it helpful to pretend that there is a semi-omniscient God, whose epistemic access to us is not different in kind or grain from the access afforded to our communities members by making marketplace observations, watching us at all times.

3.4 Summing Up

Look at one of the Sunk Cost cases. It very well may be that, at time t_2 (after sinking some costs), there's nothing more you can do to prevent suffering diachronic misfortune — outcome A , which is no longer available to you, might be better than both A^- and B . Nevertheless, at least on some occasions, you can bring about outcome B without making it *clear* that you've suffered diachronic misfortune. You can do this by, either, (1) disguising outcome B as outcome B^+ or (2) by hiding that, by your lights, A is better than B .

The outcome you end-up in is a function of both *what you've done* and *how the world is*. If the way the world is (that's relevant to your decision) is *non-public* — like, for example, if the decision turns on facts about *how you feel* — then you can, as a matter of fact, bring about outcome B but leave open the possibility of telling a plausible story about the situation according to which you've brought about B^+ instead. Alternatively, when the facts about the world on which your decision turns are public, it might yet be that how your preferences are with respect to B and A can, at least sometimes, be disguised. It is obvious that A^- is worse than A . And, while it might be true, it might not be *obvious* that B is worse than A . Recall, A^- is just like A in almost all relevant respects — except that you are poorer in A^- than in A ; and it is, *ceteris paribus*, worse to be poorer than richer. But, B is *not* just like A in almost all relevant respects. At least in some cases, then, there is a plausible story about you, consistent with your behavior, according to which you've all along preferred outcome B to outcome A .

So, that's the long and short of it. In some cases, the honoring of sunk costs

can help you disguise the fact that you've suffered diachronic misfortune. And these are the cases in which we feel *pressure* to honor sunk costs.

Is it irrational to succumb to this pressure?

4 Why is it Supposedly Irrational to Honor Sunk Costs?

Here's a first-pass at what's perhaps the line-of-thought behind the familiar admonishments of *sunk cost honoring*:

It is irrational to ϕ if there is some other available act ψ that you prefer. And by honoring sunk costs, you decide to ϕ rather than ψ , but are disposed, had sunk costs not been sunk, to ψ rather than ϕ ; and the fact that you are so disposed, *reveals* that you, in fact, *really* prefer ψ -ing to ϕ -ing — even though your actual behavior suggests otherwise.

This isn't right. You *don't* prefer *staying home* to *going to the opera*. (Of course, were sunk costs not sunk, you would prefer *staying* to *going* — but, at the very least, much more needs to be said about why this counterfactual is at all relevant). The outcomes in the actual case and the counterfactual case are different. By honoring sunk costs, then, you have not acted against your preferences.²⁰

It's not that honoring sunk costs is irrational because it amounts to acting against your preferences. Here's another suggestion. The irrationality of honoring sunk costs isn't to be found in your *action* but, rather, in your *preferences*. The problem isn't that you did something (namely, *go to the opera*) in spite of not wanting to do it — rather, the problem is: given that you'd prefer to *stay home* rather than *go* were sunk costs not sunk, it's not reasonable to prefer *going* to *staying* in the situation in which sunk costs are sunk.²¹

²⁰One thing that should be noted. I am here — and elsewhere (and everywhere) in the paper — presupposing (what could be called) **Consequentialism about Rationality**: the rationality-properties of an agent's *actions* supervene on the rationality-properties of the agent's attitudes regarding the possible *outcomes* that the agent believes will result from performing the actions. In other words, whether or not it's *rational* for someone *S* to ϕ depends on (and only on) *S*'s attitudes (e.g., her preferences) regarding the outcomes in which *S* believes ϕ -ing will result. The outcomes — or, *the consequences* — are what matters.

I intend this presupposition to be understood in the thinnest, least-controversial sense possible. That is to say: if you take yourself to have unearthed a counterexample to **Consequentialism about Rationality**, you should go on to think (as I will) that you've misspecified the relevant outcomes.

²¹You might think that the Business Major must be presupposing some **Anti-Humean** view about the rationality of preferences. After all, if you're a **Humean** about the rationality of preferences, you think that preference-profiles cannot be the object of rational

We can understand this suggestion as of a *challenge to be met*. The onus is us, the Sunk Cost Honorers, to say what it is that makes **A Night at the Opera?** relevantly different from **Short-Changed at the Opera Ticket** such that it is reasonable to prefer *going* in the former and *staying* in the latter. In other words: the task is to find a difference between the cases that is *rationally relevant*.

So far we've gone only part of the way. The quality that makes the difference, according to me, is *there being an asymmetry in the prospects for maintaining plausible deniability about suffering diachronic misfortune*. There is such an asymmetry between the options available to you in the cases in which we feel *pressure* to honor sunk costs. There is no such asymmetry, however, in those cases in which we don't feel the pressure to honor sunk costs.

If you want to be able to hide your diachronic misfortunes, you thereby have reason to honor sunk costs. Of course, if you want to poke yourself in the eye, there's at least some sense in which you thereby have reason to poke yourself in the eye. And one might think: it's *not* reasonable to poke yourself in the eye, *even if* you want to — because wanting to poke yourself in the eye is a silly and unreasonable thing to want. For any utterly bizarre behavior you can think of, we can cook up *some* desire or other such that having that desire would, at least in some sense, rationalize the behavior.

We've succeeded in pushing the challenge back a step: we've said what it is that makes the difference, but why think that this is a difference it is reasonable to let your decisions turn on? That is: why think it is reasonable to want to hide your diachronic failings? Its to answering this challenge that we will now turn.

evaluation. (“ ’tis not unreasonable for me to prefer the destruction of the whole world to the scratching of my finger” and all that). You might think: insofar as the Business Major's argument turns on criticizing the preference-profile of the Sunk Cost Honorer, it must be presupposing some sort of **Anti-Humean** view.

Maybe. But I don't think that's exactly right. As far as I understand the position, **Humeanism (about the rationality of preferences)** really only says that one's *non-instrumental preferences* are not rationally evaluable. [To put this in the parlance of Utility Theory, there is no rational difference between all the various different (consistent!) rankings of complete-world-histories (or, possible worlds). You are not rationally criticizable for the way you rank individual possible worlds.] Because our *instrumental preferences* are (or should be) sensitive to both (i) what we care about and (ii) how we take the world to be, even a **Humean** can grant that these sorts of preferences are rationally evaluable. Furthermore, it seems implausible to me — or, at least, deeply unsatisfying — that us Sunk Cost Honorers just are such that we **non-instrumentally** prefer B to A^- . The “pull” of sunk costs is a pervasive phenomena that infects all sorts of decision-problems. This suggests that there is some unifying feature — some property that all Sunk Cost cases have in common — that accounts for a whole range of Sunk Cost preferences. If you agree, then, even if you're as hardcore a **Humean** as they come, you won't be happy responding to the Business Major by leaning too heavily on your I-don't-have-to-justify-my-preference-to-anyone laurels.

5 Caring About Spinning Your Social Story

I've argued that if you want to be able to tell a plausible story about yourself that casts you in a flattering light — as someone who hasn't suffered diachronic misfortune — it is reasonable for you to honor sunk costs, when you feel the pressure to do so. In this section, I will argue that, as a matter of fact, a lot of us *do* want to be able to tell such stories about ourselves — and that this is something it is *reasonable* to expect creatures like us to want. This gives us some reason to think, then, that this isn't an irrational desire.

One way to persuasively justify the reasonableness of a want is to argue that the object of the want is a *means* to a universally-agreed-to-be-worthwhile *end*. But because we can vary the means to the ends, we have only given a weak (or, partial) rationalization. If you continue to want the means in situations where it is no longer a means to that particular end, then the want (at the very least, in those cases) is unreasonable.

It is, in some ways, more difficult to offer a persuasive justification of the reasonableness of a non-instrumental want. We can appeal to intuitions. We can, in a Humean fashion, claim that any non-instrumental want (just so long as fits in coherently with the rest of your wants) is not unreasonable — either because they are all reasonable, or 'reason' doesn't apply here at all. Justifications bottom-out somewhere. And, by definition, the usual kind of justification given for a particular want — the kind that appeals to the instrumental value of the object of the want — cannot be given when you want something *just for its own sake*.

Here's what's going to happen. Rather than either (i) offer an instrumental justification, or (ii) claim that this is a non-instrumental want and say nothing more, this is what I am going to do:

1. Claim that we want to *maintain plausible deniability about having made a diachronic mistake* non-instrumentally. This kind of self-flattering storytelling is something we can't help but want to do.
2. Offer a *Teleological Justification*: Argue that, because of the kinds of creatures we are, it was integral to our success (at achieving other ends) that we come to care about hiding our diachronic mistakes. Those of us who internalized this desire were more traditionally successful than those who didn't — and, so, through a process of Social Evolution, we've come to have this non-instrumental want.

Here's an analogy. I have, as I'm sure you do too, a *pro tanto* desire for things that taste *sweet*. When pushed, I cannot offer a satisfying justification of the reasonableness of this desire. I don't, for example, desire sweetness as the means to some end. I like things that taste sweet. I'm hard pressed to say much more than that. It isn't, though, *mysterious* why I, and creatures

like me, desire things that taste sweet. Most things that are sweet contain sugar. And sugar has fitness-promoting caloric properties. Creatures who desired sweet things did better than creatures who didn't. Even though NutraSweet doesn't contain the fitness-promoting caloric properties of sugar, it still tastes sweet to me. And even though (granting the evolutionary story I've sketched) the reason, in some sense, that I non-instrumentally desire sweetness has to do with the caloric properties of sugar, it isn't unreasonable to desire NutraSweet. I think, in some important respects, our desire to maintain plausible deniability about having suffered diachronic misfortune is like my *pro tanto* desire for sweet foods.

Here's the central claim of this section.

Claim II: It is reasonable to want *non-instrumentally* to act so as to make your diachronic actions consistent with the telling of a plausible story according to which you are diachronically virtuous (i.e., by being the kind of person that avoids making diachronic mistakes.)

The strategy for convincing you of this is to tell you a story about how having this want is good for us, and why, then, it is plausible (given our limitations) that we'd come to internalize this want.

This is the basic idea. We need to cooperate. Our success depends on it. I want you to believe, then, that I am a good cooperater. You, likewise, want me to believe this of you. Good cooperators — for various reasons — don't make diachronic mistakes. If I am to work with you, I want you to be someone (i) whose behavior is more-or-less easy to predict, (ii) who is good at predicting what will happen. If you are unpredictable, or bad at making predictions, I *ceteris paribus* will be less inclined to want you on my team.²²

We want to be able to tell flattering, plausible stories about ourselves. This is something we want non-instrumentally. It is a part of the kind of creatures that we are. We can't help but want it. Moreover, this is what we should suspect — given the kind of *deeply social* creatures that we are.

²²Here's an aside. **Our Limitations?:** perhaps, if we were Ideal Agents, we would be able to internalize desires that were more nuanced. We could, perhaps, desire to *look like a good cooperater when others are looking* — but, when others are away, our desire to tell flattering stories about ourselves subside. Point One: just as a matter of contingent psychological fact, we can't internalize and operationalize rules like that (Rule: "hide your mistakes when, and only when, others are looking). Point Two: (I'm skeptical that even the Ideal Agents would be able to do this. It would involve in engaging in some sort of delusion. You would have to think: ... "I am 'predictable' only when others watch, but others — in virtue of the fact that I think they are predictable — must actually be trying to stabilize their behavior". This is to make an epistemic exception of yourself. It seems to me to be epistemically unreasonable.

THE TELEOLOGICAL JUSTIFICATION (IN A NUTSHELL)

Social coordination is essential to our success as a social creature. Social coordination requires that I take you to be, and you take me to be, a good cooperater. In order to make myself appear like a good cooperater, I must present myself in a good light. (And, remember, being a good cooperater is essential to my success.)

Communities of Good Cooperators do better than communities of Bad Cooperators. So, we've come to internalize the desire to present ourselves in a good light. This is a non-instrumental desire because Social Evolution is unable to paint with a fine-enough brush.

5.1 Putting All the Pieces Together

Let's put the two claims, defended above, together. In order to meet the Business Major's Challenge, we need to conclude that *at least in some cases, it is reasonable to honor sunk costs.*

From the following two facts, we get our conclusion.

Claim I: The cases in which there is pressure to honor sunk costs are all such that (1) there's no plausible story to be told about you according to which you (a) sink some costs into a project, (b) later, abandon that project, and (c) don't suffer diachronic misfortune; but (2) if you carry on with the project, it is possible to tell a plausible story like this.

Claim II: It is reasonable to want non-instrumentally to act so as to make your diachronic actions consistent with the telling of a plausible story according to which you don't suffer diachronic misfortune.

It follows that it is reasonable for you to honor sunk costs in those cases in which you feel the pressure to do so. By honoring sunk costs, in these cases, you can maintain plausible deniability about having suffered diachronic misfortune. Moreover, it is reasonable to want (non-instrumentally) to maintain plausible deniability about having suffered diachronic misfortune. So it's reasonable to want to honor sunk costs. We now have something to say to the Business Major.²³

²³Are, then, the Business and Economics introductory textbooks just *wrong*? Not

6 Conclusion

Sometimes it is reasonable to honor sunk costs. Why? It's reasonable to want to maintain plausible deniability about having suffered diachronic misfortune. Sometimes, honoring sunk costs is the only way to do this. It's reasonable to want to maintain plausible deniability because having this desire is instrumental in successful cooperation, and successful cooperation is essential to our success as social creatures.

References

- [1] Arntzenius, Frank, Elga, Adam, and Hawthorne, John. (2004) "Bayesianism, Infinite Decisions, and Binding" *Mind*. 113. 450: 251-283.
- [2] Arkes, H.R. and Blumer, C. (1985) "The Psychology of Sunk Cost", *Organizational Behavior and Human Decision Processes* 35: 1224-140.
- [3] Dennett, Daniel C. (1992) "The Self as a Center of Narrative Gravity." In: F. Kessel, P. Cole and D. Johnson (eds.) *Self and Consciousness: Multiple Perspectives*. Hillsdale, NJ: Erlbaum.
- [4] Elga, Adam. (2010) "Subjective Probabilities Should be Sharp." *Philosophical Imprint*. Vol. 10, no. 5.
- [5] Kahneman, Daniel and Tversky, Amos. (1979) "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*. Vol. 47, No. 2: 263-292
- [6] Kelly, Tom. (2004) "Sunk Costs, Rationality, and Acting for the Sake of the Past." *Nous*. Vol. 38, issue 1: 60-85.
- [7] List, Christian and Pettit, Philip. (2011) *Group Agency*. Oxford University Press.
- [8] Resnik, Michael D. (2002) *Choices: an introduction to decision theory*. University of Minnesota Press.
- [9] Pettit, Philip. "Groups with Minds of their Own." *Social Epistemology: essential readings*. Oxford University Press.
- [10] Ross, Don. (2005) *Economic Theory and Cognitive Science*. MIT press.

entirely. It is a presupposition of these texts — one that is more-or-less explicit — that we're working with a specifically circumscribed set of desires: namely, the desire to amass wealth (narrowly construed). These textbooks aim to teach us how to make decisions *qua businessman* (or some such thing); not *qua human*. And, if your primary desire is to make as much money as possible, then honoring sunk costs *is* irrational. It's only when we add to the mix the desire to tell diachronically flattering autobiographies that sunk cost honoring is rational.

- [11] Schick, Frederic. (1991) *Understanding Action: an essay on reasons*. Cambridge University Press.
- [12] Stalnaker, Robert. (2002) "Epistemic Consequentialism II" *Proceedings in the Aristotelian Society Supplement*. vol. 76.
- [13] Steele, David Ramsey. (1996) "Nozick on Sunk Costs", *Ethics* 106: 605-620.
- [14] Tversky, Amos and Richard Thaler. (2004) "Anomalies: Preference Reversals." *Preference, Belief, and Similarity*. Ed, Eldar Shafir. MIT Press.
- [15] Tversky, Amos, Eldar Shafir, and Itamar Simonson. (2004) "Reason-Based Choice." *Preference, Belief, and Similarity*. Ed, Eldar Shafir. MIT Press.
- [16] Tversky, Amos. (2004) *Preference, Belief, and Similarity*. Ed, Eldar Shafir. MIT Press.
- [17] Velleman, J. David. (2005) "The Self as Narrator." In *Autonomy and the Challenges to Liberalism: New Essays*. Cambridge: Cambridge University Press.
- [18] Velleman, J. David. "Well-Being and Time." *Possibility of Practical Reason*. Oxford University Press.