The Sunk Cost “Fallacy” is Not a Fallacy

RYAN DOODY

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.

Economists and Business Majors notwithstanding, most of us do commit the sunk cost fallacy. 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 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

1For example: [McKenzie and Lee, 2006], [Frank and Bernake, 2006], and [Mankiw, 2004].

2For a collection of psychological studies to this effect, see [Arkes and Blumer, 1985], [Garland, 1990], [Moon, 2001], [Staw and Hoang, 1995]. 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 effect on one’s propensity to commit it. See [Arkes and Blumer, 1985] and [Tan and Yates, 1995], however, for some evidence to the contrary.
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 causalities) 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, 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 an endeavor we feel pressure to carry on with it. 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. Why? 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 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 hide the fact that we’ve suffered, what I will call, diachronic misfortune. Honoring your sunk costs sometimes allows you to tell a more flattering story to yourself about your diachronic behavior. And, I will argue, the desire to maintain plausible deniability about having suffered diachronic misfortune — that is, 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 a desire to tell exonerating stories about ourselves. If this is right, then honoring

---

3This isn’t just to say that we feel pressure to do so which is ultimately outweighed by other considerations. There are cases in which there is no temptation to honor sunk costs whatsoever.

4Very roughly: you suffer diachronic misfortune whenever you perform a sequence of actions that results in an outcome that is worse, by your lights, than some other outcome that could’ve resulted had you performed a different sequence of actions.
sunk costs (at least in those cases in which we feel the pressure to do so), involves satisfying a desire central to our practical identities as social creatures. And, so long as this desire is not outweighed by other considerations, it needn’t be irrational to honor sunk costs.

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 sunk costs when carrying on with the Sunk Cost Project can be better integrated into a flattering yet plausible autobiographical story than abandoning the project can be. 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.\(^5\)

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 with the following 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?

\(^5\)Is 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). I’ll hold off on fully addressing this worry until §6.
Let me make the story a bit clearer by representing it with a tree-diagram.

![Tree-diagram of A Night at the Opera?](image)

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

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.

![Tree-diagram of the Counterfactual Case.](image)

**Figure 2:** Tree-diagram of the Counterfactual Case.

**Sunk Costs:** You *honor sunk costs* if you decide to \( \phi \) rather than \( \psi \), but *holding all else fixed* are disposed, had sunk costs not been sunk, to \( \psi \) rather than \( \phi \).\(^6\)

\(^6\)One might worry that this characterization of what it is to honor sunk costs is much too broad. Suppose, for example, you know that you are psychologically wired to feel a significant amount of guilt if you don’t use the opera ticket. Because you’d rather not feel guilty, you decide to go to the opera. You aren’t honoring sunk costs; rather, you’re attempting to avoid some (future) emotional unpleasantness. And had you not bought the ticket, you wouldn’t anticipate feeling guilty about not going. So isn’t this a counterexample to *Sunk Costs*? No: *holding all*
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 *A Night at the Opera?*, I would feel considerable pressure to go to the opera. In *Short-Changed at the Opera*, I wouldn’t. 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

---

*else fixed* — in particular, that you expect to experience emotional unpleasantness if you don’t *go* — you are (presumably) *not* disposed, had sunk costs not been sunk, to *stay* home rather than go. So, it’s not obvious that this characterization is too broad. (I’ll address this worry at greater length in §6).
opera-going project *intentionally* — the opera ticket was acquired *on purpose* in *A Night at the Opera?* and acquired *accidentally* in *Short-Changed at the Opera.*

The difference between acquiring the ticket intentionally and acquiring it accidentally suggests another proposal about why we feel pressure to honor sunk costs in the former case but not the latter: by acquiring the ticket intentionally, one might think you thereby also formed the *intention* to go to the opera Saturday night; and, in general, there’s rational pressure to follow through on our intentions. But one can acquire an opera ticket accidentally without thereby forming the intention to go, and thus opt to stay home without violating any previously formed intention.7

I don’t think that this proposal is quite right either. When you purchased the ticket at $t_1$, you needn’t have formed a future-directed intention to go to the opera on Saturday. In order for it to be rational for you to form such an intention, it better be that you preferred for future-you to go to the opera over future-you staying at home. But your decision to purchase the ticket can be rational even if you lacked a preference of this sort. 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

---

7Some philosophers, see [Davidson, 1963] and [Anscombe, 1963] for example, hold that there is a tight connection between acting intentionally and acting on an intention. I intend to remain neutral about this. All I’m suggesting is that in the case at hand, your decision to purchase the ticket is accompanied by the forming of a future-directed intention to go to the opera. For a fuller discussion of intentions, and the role they play in moral psychology, see [Bratman, 2012, 2010, 1987] and [Holton, 2009].
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.³

![A Night at the Opera? (Betting)](image)

Figure 3: Tree-diagram of **A Night at the Opera? (Betting)**. Let $F$ stand for “I feel like going.”

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 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

³The 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.

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 in **Binding** to go to the opera is unconditional is a *relative* one. 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. See [Korsgaard, 2009, p. 73] on the distinction between
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 exercise if they so choose — rather, they want their future-selves to go exercise whether they feel like it at the time or not.

Sometimes our preferences are like those described in **Binding**, but at least just as often our preferences are 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. What does intentionally exchanging $100 for an opera ticket reveal about your beliefs and preferences?

Buying the ticket, as opposed to acquiring it by accident, reveals a preference, at time $t_1$, for buying over not buying. 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. The outcome that will result from your decision at 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. But, of course, acquiring the ticket accidentally reveals nothing about what your beliefs and preferences were like at time $t_1$.

Here’s another suggestion. When Saturday night rolls around, you have only 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, of course, it was accessible to you. Let me introduce some terminology:

An outcome $O$ is *diachronically accessible to you*, at a time $t_i$, if 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.

---

*bind* your future-self.\(^\text{10}\) 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 exercise if they so choose — rather, they want their future-selves to go exercise whether they feel like it at the time or not.

For some interesting discussion of self-binding in decision theory and economics, see [Arntzenius et al., 2004], [Meacham, 2010], [Elga, 2010], and [Elster, 2000].

Choosing 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 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. [Ramsey, 1926]. This idea is the cornerstone of Bayesian Decision Theory.

Exactly what the purchasing of the ticket reveals about you depends on the case. In **Binding**, in order for the purchase to be rational, you-at-$t_1$ must prefer outcome $B$ to outcome $A$ to outcome $A$. In **Betting**, where your preference are conditional on how you will feel, your decision to purchase the opera ticket is rational only if you think it reasonably likely that on Saturday you will feel like going to the opera.
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.$^{13}$

**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 $O$.

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.

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.$^{14}$ Here is an example.

**Camping Rainstorm.** 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. But what’s the difference between (i)

---

$^{13}$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).

$^{14}$It’s worth pointing out that this proposal fails for an additional reason. In *Betting*, you at all times prefer outcome $A$ to both outcome $A^-$ and outcome $B$, so no matter what you decide to do at time $t_2$, you will suffer diachronic misfortune. In fact, any time you take a bet (broadly construed) and lose, you are guaranteed to suffer diachronic misfortune.
buying a ticket, learning that you don’t feel like going, and going to the opera anyway; and, in the second case, (ii) renting 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 them. Why do we feel pressure to honor sunk costs in some cases but not others?

Here’s my hypothesis. The cases in which such pressure is felt are cases 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

---

15 There are cases in between, too. 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 is decisively drowned-out by ambient noise — in other words: cases in which we have some reason to 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.
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.

**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.

This is the idea. If you’ve suffered diachronic misfortune, then, either, (i) you’ve lost a bet or (ii) you have diachronically unstable preferences. A story in which you suffer diachronic misfortune, then, is a story according to which not everything is going your way. Weakness is unbecoming. If it is obvious that you’ve lost a bet or that you have fickle preferences, you reveal weakness. 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.

3.1 A Night at the Opera? Binding and Betting

In both versions, 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, in both versions, if you opt to go, 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 you, 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 about outcome $A$), it is possible for your to disguise the

---

$^{16}$Betting is an example of the former; Binding is an example of the latter.
preference change by going ahead with the Opera Project you 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 suffer diachronic misfortune — B is worse (and clearly so) than A. It’s 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 everything is going your way. It’s true that your action now cannot make it any less true that your preferences have changed, or that your prediction didn’t pan out — but, by deciding to go, you can effectively hide these things.\(^{17}\)

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. 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.

### 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

\(^{17}\)The 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 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 \( A^- \).

In contrast, the discussion of sunk costs in [Kelly, 2004] focuses on the potential redemptive powers that our current decisions may have on past losses. You might honor sunk costs because you desire that past sacrifices “causally contribute to that valuable state of affairs in the pursuit of which those sacrifices were originally made.” (p. 27). A desire like this would explain the pressure to honor sunk costs in cases like Binding, but it’s unclear to me that it explains the pressure to honor sunk costs in cases like Betting.
about having suffered diachronic misfortune.

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. 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 public. So, there is no plausible story about your behavior in which it rains. And, second, people just 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.\footnote{Think 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.}

\subsection*{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 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 $O$ 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

\footnote{If 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!”}

13
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 the members of our community in making marketplace observations, watching us at all times. Of course, we aren’t literally always the object of public scrutiny; although, insofar as we are both the authors of and the audience to our own behavior, there is some sense in which we are always being watched.20

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 against sunk cost honoring:

It is irrational to $\phi$ if there is some other available act $\psi$ that you prefer. And by honoring sunk costs, you decide to $\phi$ rather than $\psi$; but are disposed, had sunk costs not been sunk, to $\psi$ rather than $\phi$; and the fact that you are so disposed, reveals that you, in fact, really prefer $\psi$-ing to $\phi$-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. How are they different? Most relevantly, for my purposes, is that the former might exhibit an asymmetry in the prospects of spinning a flattering yet plausible story about your diachronic behavior. In general, we must take to individuate outcomes so as to reflect all of the relevant features that the agent cares about.21 By

20 For a discussion of the idea is that we can adopt an “outsider’s” view of our own actions, and evaluate our behaviors and ourselves from this perspective see [Smith, 1759], [Hogan and Briggs, 1986].

21 One thing that should be noted. I am here — and elsewhere (and everywhere) in the paper — presupposing 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 $S$ to $\phi$ depends on (and only on) $S$’s attitudes (e.g., her preferences) regarding the outcomes in which $S$ believes $\phi$-ing would result. The outcomes — or, the consequences — are what matters.
honoring sunk costs, then, you needn’t have acted 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* themselves. 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 this: 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.

We can understand this suggestion as a challenge to be met. The onus is us, the Sunk Cost Honorers, to find a difference between the cases that is rationally relevant. So far we’ve gone only part of the way. The feature which makes a rational difference, according to me, is the prospects for maintaining plausible deniability about suffering diachronic misfortune. There is a difference between the options available to you in the cases in which we feel pressure to honor sunk costs and the options available to you in the cases in which we don’t feel this pressure.

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,

---

22 This appears to be Kelly’s interpretation: “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.” [Kelly, 2004]

23 Does this presuppose an 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 evaluation.

Maybe. But I don’t think that’s exactly right. Humeanism (about the rationality of preferences) 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. 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 accept 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 happen to *non-instrumentally* prefer $B$ to $A^-$. The “pull” of sunk costs is a pervasive phenomenon that infects all sorts of decision-problems, only a small fraction of which involving opera tickets. 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.
we can cook up some desire or other such that having that desire would, at least in some sense, rationalize the behavior.\textsuperscript{24}

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 misfortune?

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 — then 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, we do want to be able to tell such stories about ourselves; and moreover that this is something it is reasonable to expect creatures like us to want, given our social natures. We’ve internalized a standing desire to construct flattering yet plausible autobiographical narratives about ourselves and our behavior as a way of getting along with one another. Furthermore, because these narratives give rise to “who we are” as people, this desire is deeply interwoven into our self-identities.\textsuperscript{25}

Justifying the Reasonableness of a Desire. One way to persuasively justify the reasonableness of a desire is to argue that the object of the desire 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 partial rationalization. If you continue to want the means in situations where it is no longer a means to that particular end, then the desire (at the very least, in those cases) is unreasonable.

It is more difficult to offer a persuasive justification of the reasonableness of a non-instrumental desire. We can appeal to intuitions. We can, in Humean fashion, claim that any non-instrumental desire, so long as fits in coherently with the rest of your desires, is not unreasonable (either because they are all reasonable, or ‘reason’ doesn’t apply here at all). Justifications bottom-out somewhere. Or, rather than search for an object-given reason, we might try to justify the reasonableness of a desire by offering a state-given reason.\textsuperscript{26} That is, rather than argue

\textsuperscript{24}For more discussion on this point, see [Broome, 1993], [Dreier, 1996], and [Pettit, 1991]. If there are no constraints placed on how outcomes can be re-individuated, the formal requirements of practical rationality (e.g., adopt transitive preferences) are far too easy to satisfy.

\textsuperscript{25}The claim here is similar in spirit to the account of agency in [Velleman, 2009], as well as the account of the self in [Dennett, 1992, 1989] and [Ross, 2005].

\textsuperscript{26}See [Parfit, 2001, 2011] for a fuller discussion of the distinction. Parfit thinks that all the state-given reasons in the world cannot rationalize an irrational desire.
that there’s something about X which makes it worthy of desiring, we could argue that there’s something beneficial about having the desire for X.

Here’s what I will do instead. Rather than (i) offer an instrumental justification, or (ii) claim that we desire to maintain plausibility about having suffered diachronic misfortune non-instrumentally and then say nothing more, I will:

1. Argue that we desire to maintain plausible deniability about having suffered diachronic misfortune 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, and continues to be, 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 internalize this non-instrumental desire.

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. As we’ll see, in some important respects, our desire to maintain plausible deniability about having suffered diachronic misfortune is like my pro tanto desire for sweet foods.

**Social Evolution and the Desire to Maintain Plausible Deniability.**

There is a fair amount of psychological evidence that we care quite strongly (albeit, not always consciously) about our self-presentation. For example, there’s work in the field of Self-Presentation which suggests that there is a tight connection between the impressions of ourselves that we try to create in others and our self-identities (See: [Baumeister, 1982], [Rosenberg, 1979], [Schlenker, 1980]); we sometimes attempt to define ourselves as having certain attributes by, first, attempting to convince others. Furthermore, evidence from social psychology sug-

---

27See the subtle discussion of social behavior in [Goffman, 1959], which analyzes social interaction as analogous to theatrical performance. Social interaction is akin to a performance in which “actors” create and manage the impressions they impart to their “audience.”
suggests that we engage in various behaviors (often unconsciously) with the purpose of demonstrating to others our social value, within the limits of plausibility.\footnote{See, for example, \cite{Baumeister1986, Kurzban2010, Leary2007, Schlenker1975, Schlenker1985, Tarvis2007, and Trivers2000}.} Although this behavior is aimed at managing how we are seen by others, there is evidence that also suggests that this “public” behavior is often internalized into our “private” self-conceptions; we come to see ourselves as (we imagine) others see us. \cite{Tice1992}. More specifically, a pervasive feature of self-presentation is engaging in behavior to make oneself appear diachronically consistent \cite{(Cialdini, 2001)}. In addition to this empirical evidence, there are more general theoretical reasons to expect social creatures to come to internalize a desire for spinning flattering autobiographical narratives. Here is a speculative social evolutionary story for why that might be.\footnote{This story shares similarities to the ones offered in \cite{Kurzban2010, Trivers2000}, and especially \cite{Ross2005}.}

Social coordination is essential to our success as social creatures. Social coordination requires that I take you to be, and you take me to be, a good cooperator. In order to make myself appear like a good cooperator, I must present myself in a good light. And communities of successful cooperators will do better than communities of unsuccessful cooperators. We should expect that those pro-social traits (broadly construed) conducive to successful cooperation will be selected for. The claim is that, for these reasons, we’ve come to internalize the capacities, dispositions, and sentiments necessary for successful cooperation.\footnote{The sketch of the idea here is a familiar one from Evolutionary Game Theory, wherein a pattern of behavior is represented in terms of a game-theoretic strategy, and then shown to be evolutionarily stable under certain conditions. See, for example \cite{Axelrod1986, Binmore1994, Elster1989, Frank1987, Gintis2000, MaynardSmith1982, Skyrms2004, Skyrms1996}, and \cite{Young1998}.}

We live in a social world in which choice-behavior is, very often, the subject of examination by others. Successful navigating through the world requires us to make sufficiently reliable predictions about our future behavior on the basis of fairly meager evidence about each other’s marketplace behavior. In other words, in order to get along with one another, we must construct rough-and-ready folk psychological theories of each other.\footnote{On this “interpretativist” view of rationality, see \cite{Ramsey1926, Davidson2008, Lewis1974}.} This, consequently, places rational pressure on us to “stabilize our diachronic agency” by presenting to each other \textit{coherent narratives} about our diachronic behavior; that is, we are under rational pressure to act in way that a competent observer could use to make fairly accurate predictions about our future choice-behavior. Our social nature, given the world we live in, gives us a reason to construct “selves” that are narratively coherent.\footnote{On the relationship between narrative, folk psychology, and the construction of “the self,”}
Success in the social world, however, requires more than just making ourselves *predictable* to each other. Maximally attractive prospective teammates, in addition to being diachronic stable, are not overly susceptible to taking losing bets. In brief, to make oneself into an attractive candidate for social collaboration, one must avoid the stench of failure.

Suffering diachronic misfortune is not an infallible indicator of irrationality, but it *is* a defeasible indicator of failure. If it’s revealed that you’ve suffered diachronic misfortune, you risk branding yourself as a suboptimal teammate. Here’s why. If you’ve suffered diachronic misfortune, you’ve either lost a bet or have diachronically unstable preferences. By evincing diachronically unstable preferences, you render yourself difficult to predict. If you are hard to predict, it will be difficult to coordinate with you. And if we can’t coordinate with you, you will make a less than ideal teammate. On the other hand, if it’s obvious that you’ve lost a bet, you risk establishing a reputation as a bad predictor. Of course, making a false prediction is not necessarily irrational. But given the meager amount of information we have about each other’s behavior, it’s not easy to assess whether someone’s decision to take a bet was rational or not. We want to collaborate with others who assess their evidence sensibly, proportion their beliefs to their evidence, and act sensibly in light of these beliefs. And as the number of gambles you lose increases, the likelier it may seem that you are failing on these fronts. You have some reason, then, to hide your losses (even if the bet you lost wasn’t one that it was irrational for you to have made).

Insofar as there is social evolutionary pressure to cooperate with one another, there is also pressure to present oneself as an attractive teammate. And acting so that your diachronic behavior can be woven into a flattering narrative is instrumental in presenting oneself in this sort of way. Therefore, it’s not unreasonable to expect social creatures to come to internalize a deeply-rooted desire to maintain plausible deniability about having suffered diachronic misfortune. And, because evolution doesn’t paint with a fine-brush, we should expect this desire to be internalized as a non-instrumental one.

**A Point of Clarification.** The social evolutionary story above highlighted the ways in which spinning a flattering yet plausible autobiographical narrative

---

33 This might seem like the unfair product of an irrational superstition. That might be right. But, nevertheless, we are sometimes judged by our successes and failures even if they are partly the result of chance. For example, dealers at casinos have been fired from their jobs after a sufficiently long streak of bad luck. [Goffman, 1967].

34 Moreover, it is especially embarrassing to reveal that you’ve made a false prediction about *yourself*: how you will feel, or how you will act, or what your preferences will be, etc. By revealing that you’ve made a bad prediction about yourself, you reveal that you aren’t predictable even to yourself. And if you aren’t predictable even to yourself, what hope is there for the rest of us?

---
is instrumental in achieving successful social interaction, and the benefits (social and otherwise) that come with it. And so, the desire to act in ways consistent with spinning such stories is a good desire to have, whatever your other aims might be, because it (often enough) aids in the satisfaction of these aims; and, thus, honoring sunk costs is (again, often enough) instrumentally rational. Although this desire is instrumentally valuable for these reasons, that’s not the argument. Rather, the claim is that, because spinning a flattering story about our behavior was (and continues to be) socially beneficial, we’ve come to internalize the desire to tell such stories. And, furthermore, our desire to spin flattering narratives about ourselves is a non-instrumental one; we’ve come to care about these stories for their own sake. And, if you desire something, it’s not unreasonable to treat it as a consideration in your decision-making.

To what extent does this story turn on our limitations? Maybe, if we were Ideally Rational Agents, we would be able to internalize desires that were more nuanced. Couldn’t we, for example, desire to look like a good cooperator only when others are looking? I’m skeptical that even a community of Ideally Rational Agents would be able to follow this more nuanced rule without undermining those features which made the rule socially beneficial in the first place. The desire to disguise that one has suffered diachronic misfortune facilitates coordination by making it easier for us to predict each other’s future behavior from each other’s past behavior. But if I know that you only desire to make it easy for the rest of us to predict your future behavior from your past behavior when you think the rest of us are looking (and, likewise, if you know that I only desire to make it easy for the rest of you to predict my future behavior from my past behavior when I think the rest of you are looking), then our ability to coordinate is compromised.

6 Is It Rational to Honor Sunk Costs?

Even granting all that has been said, one might still worry that it is not rational to honor sunk costs per se. It might be rational to maintain plausible deniability about having suffered diachronic misfortune by, e.g., going to the opera after having purchased a ticket, but, one might worry, that just shows that you aren’t really honoring sunk costs! In other words, if you go to the opera after buying the ticket in order to satisfy your desire to maintain plausible deniability about suffering diachronic misfortune, then your reason for going to the opera is the satisfaction of that desire and not the sheer fact that you’ve already sunk resources into the project.\(^{35}\)

Here’s an example to bring out the worry. Imagine that we live in a world in which opera tickets are wired to explode if they go unused. Given that I’d rather

\(^{35}\)See [Kelly, 2004] for a subtle treatment of this worry. [Steele, 1996] makes a similar point.
not die in an explosion, it’s completely rational of me to go to the opera after having purchased an opera ticket. However, this surely doesn’t show that it is rational to *honor sunk costs*. My reason for going to the opera, in this case, is to avoid a gruesome death, and not that I’ve already spent money on the ticket.\(^{36}\)

Isn’t something similar true about the account I’ve offered here?

Yes, but I think there is an important disanalogy. The link between maintaining plausible deniability about having suffered diachronic misfortune and honoring sunk costs is very tight. They are not merely causally correlated (given how the world is); rather, there is something closer to a *constitutive* connection between the two. In those cases in which we feel pressure to honor sunk costs, it’s *the fact that you sunk costs* which, thereby, makes it the case that your aim of maintaining plausible deniability about suffering diachronic misfortune is furthered by opting for the sunk cost option. It’s not hard to imagine a situation in which you’ve sunk costs into a ticket for the opera, and the ticket is not rigged to explode; but it is impossible to imagine a situation in which costs have been sunk into the opera ticket, but opting to forego the opera doesn’t broadcast that you’ve suffered diachronic misfortune.

Nevertheless, if one understands what it is to honor sunk costs in its *narrowest* sense, then even if these two are tightly connected (and so much so that they cannot even come apart counterfactually), one can still insist that they are *distinct* motivations. I think the right thing to say here, following [Kelly, 2004], is this. If “honoring sunk costs” is understood narrowly, then, while it’s true that nothing I’ve said here should convince you that honoring sunk costs isn’t irrational, it’s much less plausible that this is something that any of us in fact ever do. All of the paradigmatic cases of honoring sunk costs are, if “honoring sunk costs” is interpreted narrowly, not really examples of honoring sunk costs at all! On the other hand, if we interpret “honoring sunk costs” more broadly (so that, for example, the paradigmatic examples of the phenomenon count as genuine examples of it), then I’ve argued that to do so is not irrational. Furthermore, one might think that in order for something to count as a “fallacy” it has to be, on the one hand, *irrational* and, on the other hand, *tempting enough to be suitably pervasive*. And so, whether you interpret “honoring sunk costs” narrowly or broadly, either way, the Sunk Cost “Fallacy” is not really a fallacy: it’s either not irrational, or not something actual people ever do.

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

\(^{36}\)One way to motivate the thought that, in this case, you aren’t really honoring sunk costs is to ask of the counterfactual situation in which the ticket is not wired to explode (or, in which you would not feel a significant amount of regret about not using the ticket, etc.) whether you would still feel pressure to go to the opera. If you would not, then you aren’t really honoring sunk costs.
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.

### 7 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


