Ergo AN OPEN ACCESS JOURNAL OF PHILOSOPHY

The Sunk Cost "Fallacy" Is Not a Fallacy

RYAN DOODY
The University of North Carolina at Chapel Hill

Business and Economics 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 you haven't 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 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* (e.g., Frank and Bernake, 2006; Mankiw, 2004; McKenzie and Lee, 2006). 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 commit the sunk cost fallacy.[1] For the sake of picking a more neutral phrase, let's follow Nozick (1993) by referring to this behavior as *honoring sunk costs*. Examples range from the mundane to the profound, from the personal to the political. Here's

1. For 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 actually 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 (Arkes and Blumer, 1985). (See, however, Tan and Yates, 1995 for evidence to the contrary.)

http://dx.doi.org/10.3998/ergo.12405314.0006.040

one: 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 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 unwinnable. There are, of course, many other examples. In each of these situations, it's hard not to think, for example, "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 in combat will have died in vain!"

There are lots of cases in which we feel pressure to honor sunk costs. But it's not true that *whenever* we've sunk 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. There is no pressure whatsoever to honor the costs you've sunk into the insurance premiums by, for example, burning your house down. 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? In order to make a case, one way or the 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, 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 always 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*. Very 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. Honoring 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 on 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 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 a 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.[2]

2. What Is It to Honor Sunk Costs?

So far I've given only a rough characterization of what it is to honor sunk costs. To fix ideas, let's consider a canonical example.

> *A Night at the Opera?* It's Saturday night. 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 tree-diagram.

2. Is this a bait-and-switch? I draw you in with the promise of rationalizing 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.

A Night at the Opera?

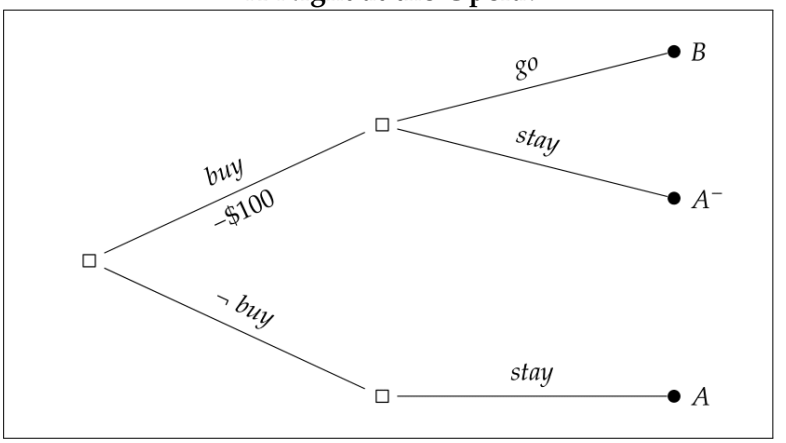

Figure 1. Tree-diagram of **A Night at the Opera?**

In cases like these, I feel pressure to go. Yet, 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.

Counterfactual Case: A Found Opera Ticket

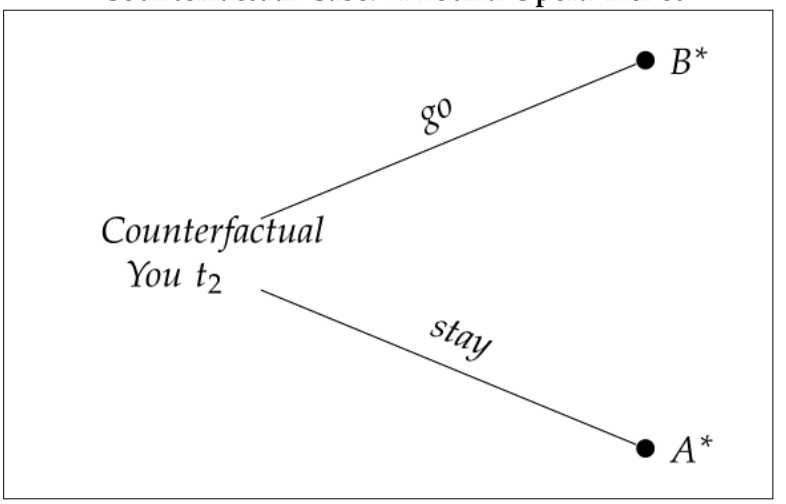

Figure 2. Tree-diagram of the Counterfactual Case.

Sunk Costs: You *honor sunk costs* if you decide to ϕ rather than ψ, but, holding all else fixed, are disposed, had sunk costs not been sunk, to ψ rather than ϕ.[3]

3. This characterization of what it is to honor sunk costs is, admittedly, rough and over-

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 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, 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), no resources have yet been expended on the opera-going project. This suggestion is not quite right, however, as the following example illustrates.

> *Short-Changed at the Opera.* You have little to no desire to go see *La Traviata* two weeks from now. 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

simplified. First, for presentational simplicity, it assumes you have only two available options. But, of course, you can honor sunk costs when you have more options, too. Second, the characterization provides only a sufficient condition for honoring sunk costs. Arguably, you also honor sunk costs when they exert *some* pressure on you to ϕ even if that pressure is ultimately outweighed by other considerations. This characterization focuses on the limiting case in which that pressure, in concert with your other reasons, is decisive. Last, and most importantly, one might worry that the characterization is much too broad (and so isn't a sufficient condition after all). Suppose, for example, that you bought a $100 ticket for the opera and now, the night of, are still excited about going. But, also suppose, that had you not bought the ticket but still wanted to go, you would need to buy a more expensive $1000 ticket at the door. Although you're happy to pay $100 to see the opera, $1000 is far too much. So, as things are, you prefer to go rather than stay but, had sunk costs not been sunk, you'd stay home. This clearly isn't an instance of honoring sunk costs. Isn't this, then, a counterexample to **Sunk Costs**? No: *Holding all else fixed*—in particular, that you don't have to dish out $1000 in order to attend the opera—you are *not* disposed, had sunk costs not been sunk, to stay rather than go. In effect, the "holding all else fixed" clause instructs you to consider a truncated decision-tree, otherwise identical to the actual one, that begins *de novo* at the current choice-node. (Honoring sunk costs, then, appears to violate the Separability Axiom of dynamic choice theory, which says, roughly, that the choice made at any node in a decision-tree should be the same as the choice that would be made in the truncated decision-tree that begins *de novo* at that node; see McClennen, 1990.) It's not obvious, then, that this characterization is too broad. However, it's also not obvious what "holding all else fixed" entails. If purchasing-and-not-using tickets reliably causes you to experience significant guilt, should we hold fixed this future emotional unpleasantness (even if you wouldn't feel guilty about not using tickets you didn't purchase)? Should we hold fixed your memories of purchasing the tickets? I'll address these worries at greater length in §6, but, for the time being, I'll trust that the intuitive idea is clear enough to continue.

> Agent—the (absolutely non-refundable-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!
>
> 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. In fact, were you to have acquired the ticket for free, it'd be a no-brainer: you'd stay home. But, alas, things aren't that simple. 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."

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. This suggests that the pressure we feel isn't owing merely 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 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 a previously formed intention.

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 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:

> **Two Versions of A Night at the Opera?**
>
> *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.) 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 the finer things. You don't want that to happen.
>
> 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 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 two weeks from now.[4]

In both versions, you might feel pressure to honor sunk costs. And while that pressure may be the result of forming an intention to go to the opera in the former version, you haven't formed an intention to go in the latter version. So the pressure to honor sunk costs doesn't perfectly coincide with the pressure to follow through on previously formed intentions.[5] 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.[6] Your buying of the ticket, in this case,

4. The decision to purchase the ticket is like taking a bet that turns on whether or not you will feel like going. (See Figure 3. F stands for "I feel like going," and $\neg F$ for "I don't 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 buy the opera ticket with the intention of going unless there's heavy snowfall that evening. (In which case, reinterpret $\neg F$ to stand for "there's heavy snowfall," etc. in Figure 3). 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.

5. Your intuitions about these cases might differ from mine. You might, for example, think that you would feel absolutely no pressure to honor sunk costs in *Betting*. That's fine. All that is required to establish the claim is that there be *some* cases in which you would feel pressure to honor sunk costs that have the same structure as *Betting*. Many classic examples of the sunk cost fallacy are more naturally represented as cases of betting than of binding. It's somewhat implausible, for example, that France and Britain preferred building the Concorde supersonic jet whether or not it would be financially sound, only to change this preference later on. The more plausible hypothesis is that they wanted to invest in the supersonic jet only insofar as it would be profitable, and were initially sufficiently confident that it would be. In other words, their decision is better thought of as a case of betting than a case of binding.

6. This isn't exactly right. It's rare that we prefer one thing to another come what may. Even

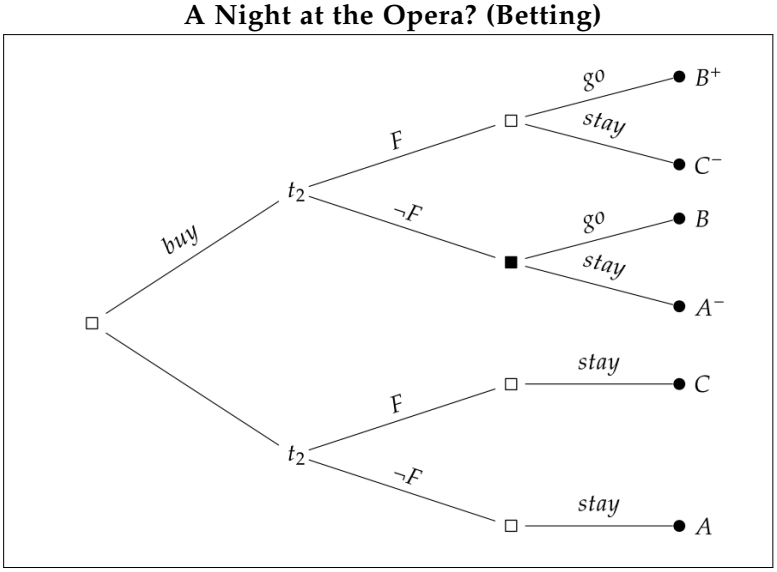

Figure 3. Tree-diagram of **A Night at the Opera? (Betting).**

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 exercise if they so choose—rather, they want their future-selves to go exercise whether they feel like it at the time or not.[7] Although sometimes our preferences are like those described in *Binding*, at least as often they 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.

---

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 might turn on. The sense then in which your desire in *Binding* 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, 73 on the distinction between treating a principle as *general* and treating a principle as *absolutely universal*.)

7. In fact, buying an expensive gym membership in order to motivate oneself to exercise more regularly is an oft-cited example of the sunk cost fallacy (McAfee et al., 2010). If you know that you're disposed to honor sunk costs and you want yourself to exercise more regularly, buying the gym membership might be a good pre-commitment strategy. Even if honoring sunk costs is irrational, this might be a sensible thing to do. It's not necessarily irrational to strategically harness future irrationality for rational ends. Nozick (1993, 22–24), for example, argues that honoring sunk costs can be rational for precisely these reasons. (See Kelly, 2004; Steele, 1996 for further discussion)

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.[8] But, of course, acquiring the ticket *accidentally* reveals nothing about what your beliefs and preferences were like at time t_1.

This naturally gives rise to 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 those times, outcome O would have resulted.[9]

8. 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 preferences 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.

9. It might be helpful to have this spelled out, slightly more formally, in terms of decision-trees. Start with a decision-tree T, consisting of choice-nodes (representing the potential moves of the agent at a particular time), chance-nodes (representing the potential moves by "nature"), and terminal nodes (representing outcomes). We can determine the outcomes that are diachronically accessible to you with the following procedure. Locate your position on T, making note of the moves of "nature" compromising your actual path from the initial node to your current position. These are the moves of "nature" that have come to pass. Hold fixed these moves by erasing from tree T those sub-trees, emanating from the chance-nodes, corresponding to the moves of "nature" that didn't come to pass. Call the resulting tree T^*. The outcomes corresponding to the terminal nodes of T^* are those outcomes that are diachronically accessible to you.

It's worth pointing out that here—but also throughout the paper—I have been implicitly assuming that the moves of "nature" are causally independent of your decisions: they would remain the same even if you had chosen differently. Although this assumption holds in some cases (e.g., your decision to purchase the opera ticket has no influence on Saturday's weather), it certainly needn't hold in all of them (e.g., your decision to purchase the ticket very well might influence what you will feel like doing on Saturday). If we turn off this assumption, the notion of diachronic accessibility sketched in this footnote needn't coincide with the one presented in the main text. The latter, roughly, looks at what would happen if you had acted differently

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. You suffer what I will call *diachronic misfortune*.

> **Misfortune** : You've suffered *diachronic misfortune* iff 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.

Notice that it takes 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 one that's diachronically accessible to you. Suffering diachronic misfortune is totally consistent with being impeccably rational.

Here's the thought: perhaps 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 not to go to the opera, you will suffer diachronic misfortune. But in *Short-Changed at the Opera*, if you decide not to go, you won't thereby suffer a misfortune of this sort.

This suggestion cannot be right, either. We sometimes do *not* feel pressure to honor sunk costs when not doing so would result in suffering diachronic misfortune.[10] 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'd feel pretty silly walking around wearing a completely ineffectual raincoat. What to do?

---

from the initial choice-node onward, allowing for the fact that "nature" might make different moves on different paths. The former, however, holds fixed the moves "nature" *actually* made and looks at how things would be if we vary your choices within the constraints given by how the world is now. Nothing in the rest of the paper will turn on this distinction, though, so let's ignore it.

10. 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.

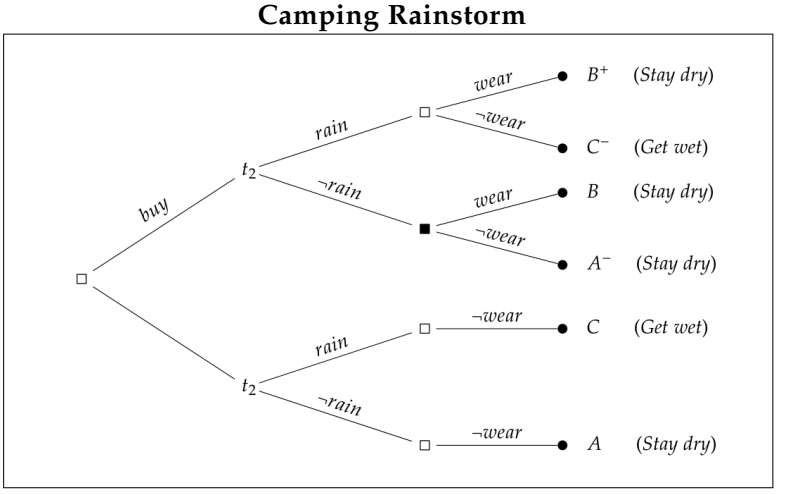

Figure 4. Tree-diagram of **Camping Rainstorm**.

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 buying a ticket, learning that you don't feel like going, and going to the opera anyway and, in the second case, 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. It is, as I will suggest in the next section, part of the story.

3. Honoring Sunk Costs and Spinning Your Social Story

There are cases in which we hear the siren call of our past expenditures luring 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.[11] 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

11. There are cases in between, too: cases in which, to stretch the already-somewhat-tired metaphor a bit more, the siren 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 in which 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 arm's 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).

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.

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, you've lost a bet or you have diachronically unstable preferences. (*Betting* is an example of the former; *Binding* is an example of the latter). 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 *those* 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—you-at-time-t_1 preferred B to A but you-now prefer A to B—and there's nothing you can do about that now. 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) and *optional* (it's not implausible for someone in your position to prefer B to A even if you in actual fact do not), it is possible for you to disguise your change in preference by going to the opera. 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 worse to do something you don't feel like doing. But, because *how you feel* is non-public (and even potentially malleable), you are able to hide the fact that you don't feel like going by opting to go.

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.[12]

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 what your preference over the relevant outcomes were or about 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.[13]

12. 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 the realization of that valuable state of affairs in the pursuit of which those sacrifices were originally made" (Kelly, 2004, 78). 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*.

13. One might worry that this isn't entirely true. If we represent the decision-problem in *Short-Changed at the Opera* so as to include your earlier decision to break your $1000 bill at the Opera Company's ticketing booth (rather than at the bank across town, or the bodega across the street, etc.), then doesn't it become clear that you've suffered diachronic misfortune in this case too? In deciding to break the bill at the ticketing booth, you took a losing bet: you hoped to get ten $100 bills and instead walked away with only nine and an opera ticket. Furthermore, can't you effectively hide that you lost this bet by opting to go to the opera? The answer, I think, depends on the extent to which you could tell a plausible story—to others, but also to *yourself*—according to which you all along wanted to pay $100 for the opera ticket.

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.

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 in which you haven't brought about a sub-optimal outcome. 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 worse than the outcome (which we've been calling A) in which you *didn't* rent the raincoat, it doesn't rain, and so you don't wear it. Just think of 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!

More importantly, if you decide to wear the raincoat anyway—despite the fact there's no rain—the prospects for telling a plausible story about your behavior in which you haven't brought about a sub-optimal outcome are also bleak. Why? Because, first, it is obvious that it isn't raining. The weather is public and non-negotiable. So there is no plausible story about your behavior in which it rains. And, second, people typically 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 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.[14] You cannot hide your diachronic misfortune by opting to wear the raincoat because it's simply not plausible—given the kinds of things that we around here typically care about—that you've all along preferred wearing the unnecessary raincoat to enjoying the sunny day having never rented the raincoat in the first place.

If your actual original aim (to break your $1000 bill, not to buy an opera ticket) is public and non-negotiable, then it's clear that you preferred the outcome in which your $1000 bill was exchanged for ten $100 bills and you stay home Saturday night to the outcome in which your $1000 bill was exchanged for nine $100 bills plus an opera ticket and you go to the opera Saturday. So, no matter what you decide to do Saturday night, it's revealed that you suffered diachronic misfortune. (In fact, your diachronic misfortune is revealed at the ticket booth, long before Saturday.) On the other hand, if your actual original aim can be obscured, going to the opera might help you maintain plausible deniability about having suffered diachronic misfortune. This depends on your ability to convince yourself that you, all along, wanted to buy the opera ticket. It's perhaps possible, in some cases, to do this. But, I claim, once we start to think of the case in this way, it's no longer obvious that we wouldn't feel any pressure to honor sunk costs.

14. 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!"

3.3. Plausible Deniability

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

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. For example, the outcome in which you've bought an opera ticket and stay home is obviously worse than the outcome in which you stay home having not bought the opera ticket. Why? Because the only relevant difference between the two outcomes is that you're $100 poorer in the former than in the latter; and it is obvious—at least, given the kinds of things that we around here care about—that you'd, all else equal, rather be $100 richer than poorer.[15]

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*. The version of *Betting* that involves predicting how you will feel is an example.

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 dis-

15. What if it were to become *common knowledge* (because new information comes to light) that the opera is terrible—that it is so bad, let us assume, that no reasonable person could prefer going to having not purchased the ticket in the first place—or that your preferences have changed? In these cases, wouldn't it be obvious that you've suffered diachronic misfortune no matter what you do Saturday night? Yes; if this were so, it would now be obvious that both of the available options are dispreferred to some diachronically accessible one. And there's some (albeit scant) empirical evidence that suggests that we wouldn't feel pressure to honor sunk costs in these cases. In a number of studies, it was found that the sunk cost effect was significantly reduced by, in various ways, making it clear between subjects and experimenters that further investments would be worse than having not invested in the first place (Berg et al., 2009; Bragger et al., 1998; Phillips et al., 1991; Tan and Yates, 1995). It certainly might seem, intuitively, like we *would* feel pressure to honor sunk costs even in these cases, but I think there's reason to be cautious here. It's easy to mistake these cases for nearby ones in which it *is* possible to disguise that one has suffered diachronic misfortune. For example, even if new information comes to light about the opera's (lack of) quality, so long as it doesn't become common knowledge that *you specifically* don't now prefer going to having not bought the ticket, it might still be possible for you to hide your diachronic misfortune by honoring sunk costs. You could, for example, tell a story (to yourself and others) about how you actually sort of enjoy terribly bad operas, or about how you enjoy watching things ironically, or about how you find the experience of going to any opera (no matter how bad) to be edifying. On the other hand, if no such story is plausible—if, for example, it becomes common knowledge that the opera performance is literally torture—then (according to my proposal) we'd no loner feel pressure to go.

guising the fact that you prefer a diachronically accessible outcome to *O*. *Binding* is an example of this, as is the version of *Betting* that involves predicting the weather.[16]

What makes a story about your behavior *plausible*? In order for a narrative to be plausible, it must be both internally and externally coherent. It's not enough that your diachronic behavior merely meets some formal constraints. The story must also attribute attitudes to you that seem reasonable to your audience. What counts as "plausible" will depend on the kinds of things that we around here—your audience—consider to be relatively natural to care about. What this amounts to, though, very much depends on your audience, their shared background knowledge of social life, and their understanding of the "social scripts" that were available to you. Plausibility can vary in degree and is sensitive to various contextual features.

The desire to maintain plausible deniability about having suffered diachronic misfortune is sensitive to our beliefs about what others believe and care about. Hiding diachronic misfortune involves disguising it from an audience—even if that audience is fictional, or hypothetical, or merely yourself—and so information about the beliefs, norms, practices, and values of your community plays an important role in delimiting what counts as "plausible."

Consider *Camping Rainstorm*, for example. There is no *plausible* story about you according to which you rent the raincoat, it doesn't rain, you wear it anyway, and you haven't stumbled into a suboptimal outcome. It's not reasonable—given the kinds of things that we around here care about—to take you to prefer wearing a rented raincoat unnecessarily to enjoying the sunny day having never rented the raincoat in the first place. People (at least, around here) don't wear raincoats on sunny days. But we can imagine a version of *Camping Rainstorm* in which your fellow campers are all members of The Society for Raincoat Appreciation: they enjoy discussing—and wearing—raincoats in all kinds of weather. In such a case, it might not be implausible that you all along preferred renting and wearing the raincoat on a sunny day to not having rented the raincoat at all; and, consequently, you might feel pressure to wear the raincoat (if you can also convince *yourself* that you have such a preference).

For whom are we constructing these narratives? Our stories are partially

16. Notice that the more hazardous the weather becomes, the less plausible it is that you prefer the outcome in which you brave the storm to see the opera over the diachronically accessible outcome in which you stay home, cozy and warm, having never bought the tickets in the first place. In the extreme—when, for example, the blizzard is so bad that only the most foolhardy would risk leaving their homes, when a State of Emergency has been issued, and spontaneous praying has broken out—it will be downright *im*plausible that you prefer going to having not bought the ticket, and so it will no longer be possible for you to disguise your diachronic misfortune. You would—and I think this is right—feel no pressure to honor sunk costs in such a case.

directed toward the other members of our 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 that of the members of our community, watching us at all times. Of course, we aren't literally the object of ceaseless 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.

3.4. Supporting the Hypothesis & Alternative Explanations

To reiterate, according to my hypothesis, we feel pressure to honor sunk costs in some cases but not others because we desire telling flattering yet plausible stories about our diachronic behavior—stories in which we haven't suffered diachronic misfortune—and honoring sunk costs can help achieve that end. Is the hypothesis true?

Although far from conclusive, there is some empirical evidence which suggests that it is. In the remainder of this section, I will present some of this evidence (and, in §5, I'll present some evidence that we *do* have such a desire) and I'll compare my proposal to some other potential explanations of why we feel pressure to honor sunk costs.

Several studies suggest that subjects have a greater propensity to honor sunk costs when they view their initial decision as a mistake for which they are personally responsible (Bazerman et al., 1982; Davis and Bobko, 1986; Staw, 1976; Staw and Fox, 1977). In all of these studies, the projects into which costs had been sunk had some (often, small) chance of ultimately being successful. This is consistent, then, with subjects investing more resources in order to (if only temporarily) disguise their misfortune. In addition, Conlon and Parks (1987) found that subjects who viewed their initial investment decision as a mistake for which they were personally responsible were significantly more likely to seek out retrospective, as opposed to prospective, information about the investment. This suggests that the subjects, who were likely to honor sunk costs, were potentially seeking ways to justify their initial investment decision (to themselves and others) as something other than a mistake. There is also evidence that subjects in these situations choose, if given the opportunity, to selectively present information that casts their decisions in a favorable light (Caldwell and O'Reilly, 1982). It's also been shown that when given the opportunity to acquire (what turns out to be) unnecessary information, subjects use that information in their decision-making in order to justify having sought it (Bastardi and Shafir, 1998). And studies of "projection bias" suggest that subjects adjust their future actions (e.g., their selling price for some object) to better align with their past predictions of those actions (Loewenstein and Adler, 1995). This evidence suggests

that we're disposed, at least under some conditions, to work toward disguising our past mistakes by taking actions now that attempt to weave them into a coherent narrative.[17] However, the evidence mentioned here is, while suggestive, far from conclusive. In particular, many (if not all) of these results are open to alternative explanations that are consistent with rival hypotheses about why we feel pressure to honor sunk costs.

There are several such rival hypotheses, but let's focus only on what I take to be the most promising three: the Avoid Waste hypothesis, the Planning hypothesis, and the Prospect Theory hypothesis. The first hypothesis, defended by Arkes and Blumer (1985), holds that the pressure to honor sunk costs derives from our desire to not appear wasteful.[18] The second hypothesis explains the pressure in terms of our dispositions to follow through on our plans, or commitments, or intentions. The final hypothesis appeals to Prospect Theory (Kahneman and Tversky, 1979)—in particular, that outcomes are evaluated as gains or losses relative to some reference point, that this reference point is subject to "framing effects," and that our value functions are convex and steep for losses—and predicts that decision-makers will honor sunk costs by carrying on with a risky project rather than settle for a sure-thing loss by abandoning it (Thaler, 1980; Whyte, 1986). I think each of these hypotheses suffer from some serious shortcomings.

Avoiding Waste. According to this hypothesis, "the avoidance of waste is a motivating factor in people's decision to honor sunk costs by not abandoning a failing course of action" (Arkes and Ayton, 1999, 595). The idea is that we desire to not appear wasteful and this desire puts pressure on us to honor sunk costs.[19]

17. It's also worth noting that, while honoring sunk costs is a pervasive phenomenon among humans, there are no (unambiguous) instances of it among lower animals (Arkes and Ayton, 1999; Curio, 1987). Arkes and Ayton (1999) also contend that young children do not honor sunk costs. If the pressure to honor sunk costs derives from a desire to construct a flattering yet plausible narrative about ourselves (which, as will be argued in §5, itself derives from our need to predict and explain each others' behavior in order to solve complex coordination problems), then, given that lower animals and young children likely lack the necessary abilities to construct such narratives, this is exactly what we should expect.

18. See Arkes (1996), Arkes and Ayton (1999) as well.

19. It's not clear that Arkes and Blumer (1985) and Arkes (1996) are making this strong of a claim, or if they are merely suggesting that there's a psychological connection *of some sort or other* between honoring sunk costs and not appearing wasteful. For the sake of argument, I'm going to address the stronger claim (and grant the weaker one). Also, Arkes and Ayton appear to endorse a slightly different hypothesis: namely, "that overgeneralization of the eminently sensible rule 'Don't waste' contributes to the manifestation of the sunk cost effect" (1999, 598). It's not fully explained how and why this rule would overgeneralize in the way that Arkes and Ayton (1999) hypothesize, but, unless the rule overgeneralizes to all and only those cases in which you're in a position to avoid appearing wasteful, we should expect this hypothesis to issue different predictions than the hypothesis in Arkes and Blumer (1985) and Arkes (1996). It's not clear from the context, however, whether this is intended to be a competing hypothesis

For example, after purchasing the opera ticket, it might appear wasteful to not use it. And so we feel pressure to go to the opera in order to avoid appearing wasteful.

Depending on how this hypothesis is spelled out, it is either unsatisfying or consistent with my own. How, in general, would honoring sunk costs help you avoid appearing wasteful? Why does it appear more wasteful to not use the tickets than it does to waste your evening at an opera that you'd otherwise prefer not to see? Moreover, we feel pressure to honor sunk costs even when there's no tangible good—like opera tickets—that we risk appearing to waste. Furthermore, we feel no pressure to wear the raincoat in *Camping Rainstorm* even though doing so would presumably appear less wasteful than renting it without using it at all.

There are answers to these worries, of course. If we appear wasteful when it's obvious that we've acquired something at a higher cost than necessary or that we've failed to efficiently use our resources, then, in order to avoid appearing this way, we might feel pressure to act so that a plausible story can be told about our behavior according to which these things are not the case. Not using the opera ticket appears more wasteful than wasting your evening because it's easier to hide your feelings about the opera than it is to hide unused opera tickets. There's no pressure to wear the raincoat because, although not wearing it might appear wasteful, it's already obvious that renting it was a waste. But notice that if you've acquired something at a higher cost than necessary or if you've failed to efficiently use your resources, you've also suffered a diachronic misfortune. So it's no longer clear, when spelled out in more detail, whether this is a *rival* hypothesis after all.

Adhere to Your Plans. According to this hypothesis, the pressure we feel to honor sunk costs derives from a generally admirable propensity to follow through with our plans. Making plans, and then following through on them, is beneficial for various reasons: for example, it helps us achieve long-term goals in the face of temptation; it allows us to avoid the costs of continually reconsidering our reasons for action by closing off future deliberation; it facilitates inter- and intra-personal coordination; etc. (Bratman, 1987). Once a plan has been made, we (at least, typically) feel psychological pressure to follow through on it. Perhaps, then, the pressure we feel to honor sunk costs is *really* pressure to follow through on a plan.

But, as we've observed (in cases like *Betting*), we feel pressure to honor sunk costs even when no such plan has been made (or, rather, when we've made a *conditional* plan such that adhering to it isn't served by taking the sunk-cost-honoring option). You bought the opera ticket, not with the plan to go, but to give yourself the option to if you feel like it. So the pressure to honor sunk

or not.

costs doesn't derive straightforwardly from the impulse to follow through on our plans. Instead, the pressure to honor sunk costs might result from an *overgeneralization* of our general propensity to follow through on our plans. Our impulse to follow through on our plans is so strong that we feel pressure to do so even when no such plan has been made.

I have a couple worries about this explanation. First, while it's not implausible that our propensity to follow through on our plans might overgeneralize to cases in which no such plans have been made, it's unclear why such a propensity would overgeneralize to those cases in which we feel pressure to honor sunk costs but not overgeneralize to those cases in which we don't. Compare *Betting* to *Camping Rainstorm*. In both cases, you don't adhere to your original (conditional) plan by honoring sunk costs: you planned to go to the opera but only if you feel like it (or only if the weather isn't terrible); you planned to wear the raincoat but only if it rains. We feel pressure to honor sunk costs in the former case but not the latter. Given the structural similarity of the two cases, why would our propensity to follow through on our plans overgeneralize to the former but not the latter?

Or imagine a case a lot like *Betting* but with no sunk costs: you have a standing invitation to see the opera, at no cost, whenever you'd like to; you make a plan to go next Saturday unless the weather is terrible; on Saturday, the weather is terrible. I wouldn't feel pressure to go to the opera in this case. Why would the pressure to make good on our plans only overgeneralize to the version of the case in which costs have been sunk?

There is also an interesting phenomenon—the so-called "Reverse Sunk Cost Effect" (Heath, 1995) or "Pro Rata Fallacy" (Baliga and Ely, 2011)—wherein decision-makers honor sunk costs by *abandoning*, rather than following through on, the project into which costs have been sunk. In such cases, upon learning that the total costs needed to successfully complete a project exceed its value, decision-makers are reluctant to continue investing additional resources into its completion (even if abandoning it is likely to result in an even greater net loss). Here's an example. Suppose you bought an old house with the plan to renovate and re-sell. After purchasing the house, though, the real-estate market takes a turn. It becomes clear that you won't be able to recoup your total expenses by renovating and re-selling. Instead, you could re-sell the house as-is (also at a loss). There's evidence that people feel pressure to abandon their original plan in favor of re-selling the house as-is even though, had sunk costs not been sunk (had they been given the house as a gift, for example), they would prefer to re-sell the house after completing the renovations.[20] This is an example of honoring sunk

20. See Heath (1995) for a number of experiments suggesting that subjects "de-escalate" investment in response to sunk costs precisely in cases like the one presented above. See also Baliga and Ely (2011), who present and experimentally test a memory-based model of the

costs—you feel pressure to abandon the project rather than carrying on with it even though, had sunk costs not been sunk, you would be disposed to carry on with the project rather than abandon it—that is difficult to explain in terms of adhering to your plans. Even if the impulse to follow through on our plans overgeneralizes, in cases like these, we feel pressure to abandon, not *follow through* on, the plan.[21]

Prospect Theory. Another explanation for why we honor sunk costs appeals to Prospect Theory (Kahneman and Tversky, 1979), which is a descriptive theory of decision-making under risk. Prospect Theory deviates from Expected Utility Theory in several respects, but the important differences, for our purposes, are these: first, outcomes are evaluated as gains or losses relative to a reference point; second, this reference point is determined by how the decision-problem is "framed" psychologically; and decision-makers are assumed to have an *S*-shaped utility curve, which kinks at the origin (the reference point), is concave for gains, and is both convex and steep for losses. In effect, Prospect Theory predicts that decision-makers will exhibit risk-inclined behavior when choosing between perceived losses and risk-averse behavior when choosing between perceived gains. The explanation holds that sinking costs into a project affects our reference point so that abandoning the project is perceived as a sure-thing loss. Because we are, according to Prospect Theory, averse to sure-thing losses, we will feel pressure to honor sunk costs (Thaler, 1980; Whyte, 1986).

There are a number of problems with this explanation as well. First, even if Prospect Theory were to correctly predict when decision-makers will, and will not, honor sunk costs, it's not obvious that it provides a suitable *explanation* of the phenomenon. As Arkes and Blumer complain, "prospect theory does not specify the psychological basis for the findings that sure losses are so aversive and sunk costs are so difficult to ignore" (1985, 132). As is, Prospect Theory

phenomenon. (I think that their model, while interesting, suffers from several theoretical and empirical problems. But discussing it further would take us too far afield.)

21. Notice, however, that my explanation *can* potentially account for these cases. If you carry on with the project, you will bring about an outcome that's clearly and obviously worse than what would've resulted had you not purchased the house in the first place. But, if you abandon the project by re-selling the house as-is, you have the opportunity to hide your diachronic misfortune by bringing about an outcome (one in which you invest your resources elsewhere) that, while perhaps also worse, looks less obviously so to an "outside observer." Moreover, Heath (1995) found that subjects are less likely to abandon a failing project when the subsequent investments are "difficult to track" (e.g., investing time when the sunk costs are money, or investing money when the sunk costs are time). When the value of the overall expenses allocated to a project is equivocal, it's easier to disguise whether completing the project would be worse than never having begun it. And so, in these cases, abandoning the project no longer affords you a better opportunity for hiding your diachronic misfortune than carrying on with it would.

provides at best a *model* of sunk cost honoring, not an explanation of it. Worse, it's not clear that Prospect Theory *does* correctly predict when sunk costs will be honored (see Schoorman et al., 1994, for example). Furthermore, the explanation crucially assumes that honoring sunk costs isn't also a sure-thing loss (relative to the reference point). That needn't be the case. You needn't think of going to the opera as a risky option—one that might, with some (perhaps low) probability, result in an outcome that you prefer to having not purchased the ticket in the first place—in order to feel pressure to honor your sunk costs. Lastly, given the structural similarities between *Betting* and *Camping Rainstorm*, it's unclear how Prospect Theory could explain why we feel pressure to go to the opera in the former case but don't feel pressure to wear the raincoat in the latter.

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 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 ψing to ϕing—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 should individuate outcomes so as to reflect all of the relevant features that the agent cares about. By 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.[22] 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.

22. This is 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, 62).

We can understand this suggestion as a challenge to be met. The onus is on us, the honorers of sunk costs, 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, 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?

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 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 with our self-identities.

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, this would constitute only 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 fash-

ion, 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.[23] That is, rather than argue 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.[24]

Here's what I will do instead. Rather than offer an instrumental justification, or 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 impor-

23. 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. One might think that state-given reasons for having a desire D, at best, rationalizes the desire to be such that you have desire D.

24. This is, roughly, the strategy Nozick (1993) adopts in justifying the honoring of sunk costs. It's good, according to Nozick (1993), that we honor sunk costs because we can strategically exploit the fact that we will honor sunk costs in order to help our future-selves overcome temptation. This is a state-given, and not an object-given, reason to honor sunk costs. Steele (1996) criticizes Nozick's argument, largely for this reason.

tant 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 empirical evidence that we quite strongly (albeit not always consciously) care about our self-presentation.[25] For example, Kurzban and Aktipis argue that we've internalized a set of mechanisms that are "designed for strategic manipulation of others' representations of one's traits, abilities, and prospects" (2007, 131). These mechanisms work to strike the optimal balance in self-presentation between favorability and plausibility (Baumeister, 1982; Schlenker, 1975) with the aim of demonstrating our social value to others.[26] One primary function of these mechanisms is to maintain the appearance of consistency (Swann, 1985; Tedeschi et al., 1971; Stone et al., 1997). And, although these mechanisms serve a social function, there's evidence that the mechanisms exert motivational force on us even in private (Baumeister, 1982; Hogan and Briggs, 1986; Schrauger and Schoeneman, 1979; Tice and Baumeister, 2001); there is a tight connection between the impressions of ourselves that we attempt to instill in others and our own self-identities (Baumeister, 1982; Kurzban and Aktipis, 2007; Rosenberg, 1979; Schlenker, 1980). Kurzban and Aktipis metaphorically likens these mechanisms to a press secretary: "[I]t collects and stores information about what one has done and engages in spin to make the individual's actions appear as positive as possible" (2007, 136). Furthermore, they argue that the motives embodied by the "press secretary"—namely, the desire to construct plausible autobiographical narratives that cast its protagonist in a favorable light—operate without conscious awareness (and, they suggest, for good reason: conscious awareness of such a desire might undermine its satisfaction; see Trivers 2000).[27]

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. One such theoretical reason is the following speculative social evolutionary story.[28]

---

25. See 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."

26. See, for example, Baumeister (1986), Kurzban (2010), Leary (2007), Schlenker (1975, 1985), Tarvis and Aaronson (2007), and Trivers (2000).

27. In order to effectively convince others, it's often helpful to first convince ourselves. But if you're consciously aware of your desire to hide your diachronic misfortune, it will be exceedingly difficult (and perhaps impossible) to convince *yourself* that you haven't suffered diachronic misfortune.

28. This story shares similarities to the ones offered in Kurzban (2010), Trivers (2000), and especially Ross (2005). (I outline this story in more, and slightly different, detail in Doody, 2019.)

Social coordination is essential to our success as social creatures (Kurzban, 2010; Levine and Kurzban, 2006; Tooby and Cosmides, 1996). 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 (Brewer, 1997; Hauser, 1996; Trivers, 2000). Because 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.

We live in a social world in which our choice-behavior is very often the subject of examination by others. Successful navigation through this world requires us to make sufficiently reliable predictions about each other's future behavior on the basis of fairly meager evidence about each other's past behavior. To get along with one another, we must construct rough-and-ready folk psychological theories of each other. This is no easy task. Consequently, we face rational pressure to stabilize our diachronic agency by presenting to each other *coherent narratives* about our diachronic behavior. We have reason to act so that a competent observer would be able to make fairly accurate predictions about our future choice-behavior on the basis of our past choice-behavior.[29]

Success in the social world, however, involves more than merely making ourselves *predictable* to one another. It also involves allying ourselves with others—prospective teammates—who are reliably successful at securing their ends. We often do better by working together than by going it alone. But, by working together, we condition our success on the success of others: our teammates. And so, it's in each of our interests to choose, and be chosen by, prospective teammates who are successful. Because teams must navigate dynamic environments, they should want their members to be *reliably* successful: to have a set of skills that are success-conducive in a wide array of situations. Assessing one's evidence well, proportioning one's beliefs to one's evidence, making sensible decisions in light of these beliefs, etc. are all examples of skills that contribute to reliable success. To earn a spot on an attractive team, you must appear like an attractive candidate. It's in your interest, then, to highlight your successes and to downplay your failures.

29. The relationship between narrative, folk psychology, and the construction of "the self" has been explored in both philosophy (Dennett, 1992, 1989; Ross, 2005; Velleman, 2005, 2009) and cognitive science (Gazzaniga, 1998; Goldie, 2012; Hutto, 2007). A common theme throughout is the importance of the role narrative plays in social coordination, which often requires presenting a unified account of our behavior. This is related to what McGeer calls "the regulative dimension of folk psychology," central to which is the claim that "skilled folk psychologists are not just able to read other people in accord with shared norms; they also *work* to make themselves readable in accord with those same norms" (2007, 148).

Suffering diachronic misfortune, while not an infallible indicator of irrationality, *is* an indicator of failure: you've failed at bringing about an optimal outcome. Because appearing reliably successful is instrumental in securing a spot on an attractive team, by revealing your diachronic misfortune, you risk damaging your reputation as a suitable teammate. Here's why. If you've suffered diachronic misfortune, then either (1) you've exhibited diachronically unstable preferences or (2) you've lost a bet.

Consider (1). By exhibiting diachronically unstable preferences, you render yourself hard to predict.[30] 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. There's pressure on us, then, to present ourselves in ways that uphold the appearance of diachronic consistency (Cialdini, 2001; Stone et al., 1997; Swann, 1985; Tedeschi et al., 1971).[31]

Consider (2). Although losing a bet is compatible with reliable success, *revealing* this loss might cause prospective teammates to form an unfavorable impression of you. It's not unreasonable for you to worry that, all else equal, they are more likely to form such an impression if you reveal your loss than if you don't. Because it is in your interest for them to not form this impression, you have reason to hide your losses when it's not difficult to do so (even though this loss may entirely be the product of bad luck). When there's competition for spots on the team, it's risky to hope that others will grant you the benefit of the doubt; it's safer to avoid, if you can, even the possibility of looking incompetent.

Of course, even perfect decision-makers are occasionally unlucky; the best choice *ex ante* needn't be the best *ex post*. Given that bad outcomes can result from good decisions, is it really reasonable to worry that revealing a bad outcome will damage your reputation as a good decision-maker? Typically, yes. Given the meager amount of information we have about each other, it's often not possible to *directly* assess the quality of each other's decision-making abilities. Were

30. Diachronically unstable choice-behavior is difficult to rationalize as the product of coherent beliefs and desires had by a unified agent who cares about things in ways that we around here find intelligible. It's not difficult, in general, to rationalize an agent's behavior if we are allowed to individuate the outcomes of the decision-problems the agent faces as finely as need be, which amounts to representing the agent's preferences as sensitive to those features individuating the outcomes (Broome, 1993; Dreier, 1996; Pettit, 1991). But we rescue the unified agent's (formal) coherence at the expense of representing her as caring about things we around here might find hard to understand. Either way, our ability to predict the agent's behavior suffers.

31. What counts as "diachronically consistent" is a more complicated matter than I'm letting on. One can suffer diachronic misfortune as the result of diachronically unstable preferences in a way that doesn't make one's future behavior hard to predict. For example, predictable preference shifts—like those that standardly occur as we mature, or like those that typically accompany significant life changes—in virtue of being predictable, needn't undermine our ability to coordinate with each other. I address this issue in more detail in Doody (2019).

we to have direct access to your mind at the time of the decision, the outcome shouldn't matter to our assessment of you as a decision-maker. We'd already know all that we'd need to know. But, because we typically have only limited information about the basis on which a decision was made (e.g., we don't know what evidence was available at the time, how well this evidence was assessed, the probabilities that were assigned to the outcomes at the time, what other considerations were considered relevant, etc.), learning that it resulted in a clearly suboptimal outcome *suggests* (albeit defeasibly) something about the quality of the decision and the competence of the decision-maker. The outcome of your decision, in absence of further information about how it was made, is evidentially relevant to your decision-making ability: insofar as we think it's more likely for suboptimal outcomes to result from poorly-made decisions than from well-made ones, suboptimal outcomes are evidence of the former. So, given the poverty of information typically available, it's understandable why we might be disposed more favorably toward bet-winners than bet-losers. But even if such favoritism is fallacious and unfair, so long as it is reasonable for you to *worry* that these are the evaluative standards in place, you have reason to highlight your wins and to downplay your losses.[32]

Moreover, it is especially embarrassing to reveal that you've lost a bet about *yourself* (concerning, e.g., how you will feel, what you will do, what your pref-

32. There is a considerable amount of empirical evidence suggesting that we in fact *do* evaluate decisions on the basis of their outcomes—and that this continues to be the case even when we are fully-informed about the basis on which the decision was made. Evaluators suffer from what Baron and Hershey (1988) call *outcome bias*, which has been observed in a wide range of domains (e.g., finance: Baron and Hershey, 1988; Germann and Weber, 2018; Konig-Kersting et al., 2017; Zakay, 1984; medicine: Baron and Hershey, 1988; politics: Wolfers, 2002; Gasper and Reeves, 2011; and sports: Lefgren et al., 2014; Kausel et al., 2018). Why might evaluators be outcome biased? Here are three possibilities. First, it could be that evaluators (mistakenly) take luck *itself* to be a hidden skill that some have more of than others (Darke and Freedman, 1997; Langer, 1975). If some people are inherently luckier than others, then outcomes provide useful information about the propensity of a decision-maker to be successful in the future. Second, outcome bias might be the product of a generally helpful but misapplied heuristic. When we lack relevant information about the basis on which the decision was made, its outcome is an imperfect indicator of its quality. Typically, we do lack this information. So it's typically rational to take outcomes into account when evaluating the quality of a decision (Hershey and Baron, 1992). The impulse to do so, however, overgeneralizes: we continue to take outcomes into account even when it is inappropriate to do so. Third, it might be that evaluators are making a holistic assessment of the decision-maker's ability to be reliably successful rather than an assessment of that specific decision's rationality. For example, even if evaluators know that a specific decision was made rationally given what the decision-maker believed at the time, they might take a suboptimal outcome to suggest that the decision-maker could've gathered better evidence prior to making the decision. Whatever the explanation, if evaluators are outcome biased and it is in your interest to be evaluated favorably, it will also be in your interest to manipulate what evaluators might learn about the outcomes of your decisions (which is what Brownback and Kuhn, 2018 found in their study).

erences will be, and the like). When making a prediction about yourself, it's presumed (perhaps, falsely) that you occupy a privileged position with respect to the relevant evidence, and it's often particularly opaque to others exactly what this evidence specifically is. The more private your evidence, the more vulnerable you are to charges that you failed to assess it correctly. And, furthermore, by revealing that you've lost such a bet, you suggest that you aren't predictable even to yourself. And, as prospective teammates might very well worry, if you aren't predictable to yourself, what hope is there for the rest of us? Someone who is bad at predicting what they themselves will do is someone for whom it's reasonable to think it will be difficult for the rest of us to predict as well.

In order to broadcast your social worth as a potential teammate, you want to appear as though your preferences are stable, you've assessed your evidence well, you've appropriately accounted for risk, and you've made sensible decisions. Because diachronic misfortune involves either unstable preferences or a lost bet, revealing that you've suffered it risks undermining your appearance as a worthwhile teammate.[33] Insofar as there is social evolutionary pressure to cooperate with one another, there is also pressure to present oneself as an attractive teammate. 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.

33. One might object that, while this might typically be true—that revealing a change in preference makes you harder to predict, and that revealing that you lost a bet can (for the reasons outlined above) be taken as a sign that you failed to sagaciously assess your evidence, proportion your beliefs to that evidence, and act sensibly in light of those beliefs—it's not always the case. For example, if it's common knowledge that, when deliberating about whether to take some bet, you assessed your evidence well and that you maximized expected utility, then revealing that you lost the bet doesn't suggest that you've made an unflattering mistake. Given that revealing diachronic misfortune doesn't *always* undermine your appearance as a worthwhile teammate, why think there's evolutionary pressure to internalize a desire to disguise your diachronic misfortune rather than pressure to internalize a desire that more closely tracks these underlying features? There are two reasons. First, as discussed in the previous footnote, if evaluators are outcome biased (and there's considerable evidence that they are), revealing your diachronic misfortune will damage your reputation even in these cases. The second reason is that these cases—ones in which the facts about your evidence-assessment and deliberation are common knowledge—are very rare in real-world situations, and so it's unlikely that the social evolutionary forces would track them. It's typically very difficult to know what someone's evidence is (much less whether they assessed it correctly), or the basis on which they made a particular decision, etc. We shouldn't expect the evolutionary forces to be able to distinguish between desires whose differences mainly manifest in such atypical cases.

A Point of Clarification. The social evolutionary story above highlighted the ways in which spinning a flattering yet plausible autobiographical narrative 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 for it to factor in your decision-making.

To what extent does this story turn on our limitations? Maybe if we were *ideally rational agents*—impeccable bayesian agents with no cognitive limitations—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?*

Perhaps, but I'm skeptical. Ideal rationality isn't *omniscience*. Because the ideally rational agent, just like us, cannot always discern when her behavior is the subject of others' scrutiny, it's not obvious that a community attempting to follow the more nuanced rule would enjoy as much cooperative success as a community following the less nuanced rule. The ideally rational agent could, on each occasion, run the cost-benefit calculations (taking into account her credence that others are looking) and decide to do whatever maximizes expected value. If the costs of being found out, and her credence that others are looking, are high enough, she will keep up appearances; otherwise, she will not worry about looking like a subpar teammate. But sometimes she'll be wrong; she'll fail to keep up appearances when others are looking, and she'll undermine her reputation as a suitable teammate as a result. Even if these occasions are rare for each individual, an entire community of ideally rational agents behaving in this way might be expected to cooperate less successfully than a community implementing the less nuanced rule.[34]

Recall that 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, and that such coordination is instru-

34. Gintis and Helbing credit the success of *Homo sapiens* to the internalization of norms, making a similar point to the one being gestured at here: "When individuals internalize a norm, the frequency of the desired behavior will be higher than if people follow the norm only instrumentally—i.e., when they perceive it to be in their best interest to do so on self-regarding grounds" (2015, 11).

mental in our collective success. 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; we can no longer predict each other's future behavior as reliably. You're entitled to infer a great deal less about how I will act in the future from what you observe about my current behavior if you know that how we act when we think we're being observed might differ from how we act when we don't. Analogously, we should expect the community that internalizes a norm prohibiting lying to cooperate more effectively than the community that internalizes a rule like "Don't lie unless you think you can get away with it." The latter community will have trouble establishing the requisite amount of trust necessary for successful social cooperation. Similarly, the community of ideally rational agents who attempt to appear like good cooperators only when others are looking will fair worse than the community that internalized the less nuanced desire.

6. Is It Rational to Honor Sunk Costs?

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, for example, 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 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]

35. See Kelly (2004) for a subtle treatment of this worry. Steele (1996) makes a similar point when distinguishing between the honoring of sunk costs and the desire to finish what we've started.

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 (but *holding all else fixed*) whether you would still feel pressure to go to the opera. If you would

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 introductory Business 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 amass wealth (narrowly construed). These textbooks aim to teach us how to make decisions *qua businesspeople* (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.

---

not, then you aren't really honoring sunk costs. It's not always obvious how to evaluate such counterfactuals, however, especially because it's not entirely clear what "holding all else fixed" involves.

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.

Acknowledgments

I've been working on versions of this paper for a long time, so I have very many people to thank. First, for inspiration, I should thank my parents (who are seemingly in an endless debate about how much weight to give to sunk costs). For helpful feedback and discussions, I am grateful to Arden Ali, Lara Buchak, Jennifer Carr, Nilanjan Das, Tom Dougherty, David Gray Grant, Dan Greco, Daniel Hagen, Toby Handfield, Sally Haslanger, Brian Hedden, Richard Holton, Sophie Horowitz, Abby Everett Jaques, Brendan de Kenessey, Cole Leahy, Jack Marley-Payne, Nicole Mazzeo, Emily McWilliams, Eleanor Muse, Sofia Ortiz-Hinojosa, Damien Rochford, Said Saillant, Miriam Schoenfield, Melissa Schumacher, Jack Spencer, Bob Stalnaker, and surely many others. I also want to thank the participants and the audiences at the 2012 Northwestern Time and Rationality Workshop, the 2013 Rocky Mountain Philosophy Conference (especially Caleb Pickard, who provided comments), the 2014 Annual Meeting of South Carolina Society of Philosophy, The 2014 Syracuse Philosophy Graduate Conference (especially Steve Steward, who provided comments), the 7th annual LSU Philosophy Conference, the 2014 Gateway Graduate Conference at the University of Missouri, St. Louis (especially Kevin Rice, who provided comments), the 2014 Columbia/NYU Philosophy Graduate Conference (especially Amanda Askell, who provided comments), the 2014 UMASS Amherst Philosophy Graduate Conference (especially Lisa Cassell, who provided comments). I owe special thanks to Caspar Hare, Agustín Rayo, Steve Yablo, and to several anonymous referees and editors.

References

Anscombe, Gertrude E. M. (1963). *Intention* (2nd ed.). Blackwell.

Arkes, Hal R. (1996). The Psychology of Waste. *Journal of Behavioral Decision Making*, *9*(3), 213–224.

Arkes, Hal R. and Peter Ayton (1999). The Sunk Cost and Concorde Effects:

Are Humans Less Rational than Lower Animals? *Psychological Bulletin*, *125*(5), 591–600.

Arkes, Hal R. and Catherine Blumer (1985). The Psychology of Sunk Costs. *Organizational Behavior and Human Decision Processes*, *35*(1), 124–140.

Baliga, Sandeep and Jeffrey C. Ely (2011). Mnemonomics: The Sunk Cost Fallacy as a Memory Kludge. *American Economic Journal: Microeconomics*, *3*(4), 35–67.

Baron, Jonathan and John C. Hershey (1988). Outcome Bias in Decision Evaluation. *Journal of Personality and Social Psychology*, *54*(4), 569–579.

Bastardi, Anthony and Eldar Shafir (1998). On the Pursuit and Misuse of Useless Information. *Journal of Personality and Social Psychology*, *75*(1), 19–32.

Baumeister, Roy F. (1982). A Self-Presentational View of Social Phenomena. *Psychological Bulletin*, *91*(1), 3–26.

Baumeister, Roy F. (Ed.) (1986). *Public Self and Private Life*. Springer.

Bazerman, Max H., Rafik I. Beekun, and F. David Schoorman (1982). Performance Evaluation in Dynamic Context: The Impact of a Prior Commitment to the Ratee. *Journal of Applied Psychology*, *67*(6), 873–876.

Berg, Joyce, John Dickhaut, and Chandra Kanodia (2009). The Role of Information Asymmetry in Escalation Phenomena: Empirical Evidence. *Journal of Economic Behavior and Organization*, *69*(2), 135–147.

Bragger, Jennifer, Donald Bragger, Donald Hantula, and Jean Kirnan (1998). Hysteresis and Uncertainty: The Effect of Uncertainty on Delays to Exit Decisions. *Organizational Behavior and Human Decision Processes*, *74*(3), 229–253.

Bratman, Michael (1987). *Intension, Plans, and Practical Reason*. Harvard University Press.

Bratman, Michael (2010). Agency, Time, and Sociality. *Proceedings and Addresses of the American Philosophical Association*, *84*(2), 7–26

Bratman, Michael (2012). Time, Rationality, and Self-Governance. *Philosophical Issues*, *22*(1), 73–88.

Brewer, Marilynn B. (1997). On the Social Origins of Human Nature. In Craig McGarty and S. Alexander Haslam (Eds.), *The Message of Social Psychology: Perspectives on Mind in Society* (54–62). Blackwell.

Broome, John (1993). Can a Humean Be Moderate? In Raymond Frey and Christopher W. Morris (Eds.), *Value, Welfare and Morality* (51–73). Cambridge University Press.

Brownback, Andy and Michael A. Kuhn (2018). Understanding Outcome Bias: Asymmetric Sophistication and Biased Beliefs. Manuscript in preparation. Available at SSRN: https://dx.doi.org/10.2139/ssrn.3059030

Caldwell, David F. and Charles A. O'Reilly (1982). Response to Failure: The Effects of Choice and Responsibility on Impression Management. *Academy of Management Review*, *25*(1), 121–136.

Cialdini, R. B. (2001). *Influence: Science and Practice* (4th ed.). Allyn and Bacon.

Conlon, Edward J. and Judi M. Parks (1987). Information Requests in the Context of Escalation. *Journal of Applied Psychology*, *72*(3), 344–350.

Curio, Eberhard (1987). Animal Decision-Making and the Concorde Fallacy. *Trends in Ecology and Evolution*, *2*(6), 148–152.

Darke, Peter R. and Jonathan L. Freedman (1997). The Belief in Good Luck Scale. *Journal of Research in Personality*, *31*(4), 486–511.

Davidson, Donald (1963). Actions, Reasons and Causes. *The Journal of Philosophy*, *60*(23), 685–700.

Davis, Mark and Philip Bobko (1986). Contextual Effects on Escalation Processes in Public Sector Decisions. *Organizational Behavior and Human Decision Processes*, *37*(1), 121–138.

Dennett, Daniel C. (1989). The Orgins of Selves. *Cogito*, *3*(3), 163–173.

Dennett, Daniel C. (1992). The Self as Center of Narrative Gravity. In Frank Kessel, Pamela M. Cole, and Dale L. Johnson (Eds.), *Self and Consciousness: Multiple Perspectives* (103–115). Lawrence Erlbaum.

Doody, Ryan (2019). If There Are No Diachronic Norms of Rationality, Why Does It Seem Like There Are? *Res Philosophica*, *96*(2), 141–173.

Dreier, James (1996). Rational Preference: Decsion Theory as Theory of Practical Rationality. *Theory and Decision*, *40*(3), 249–276.

Frank, Robert and Ben Bernake (2006). *Principles of Microeconomics* (3rd ed.). McGraw-Hill.

Garland, Howard (1990). Throwing Good Money after Bad: The Effect of Sunk Costs on the Decision to Escalate Commitment to an Ongoing Project. *Journal of Applied Psychology*, *75*(6), 728–731.

Gasper, John T. and Andrew Reeves (2011). Make It Rain? Retrospection and the Attentive Electorate in the Context of Natural Disasters. *American Journal of Political Science*, *55*(2), 340–355.

Gazzaniga, Michael S. (1998). *The Mind's Past*. University of California Press.

Germann, Maximilian and Martin Weber (2018). Outcome Bias in Financial Decision Making. Manuscript in preparation. Available at SSRN: https://dx.doi.org/10.2139/ssrn.3247148

Gintis, Herbert and Dirk Helbing (2015). Homo Socialis: An Analytical Core for Sociological Theory. *Review of Behavioral Economics*, *2*(1-2), 1–59.

Goffman, Erving (1959). *The Presentation of Self in Everyday Life*. Anchor.

Goldie, Peter (2012). The Narrative Sense of Self. *Journal of Evaluation in Clinical Practice*, *18*(5), 1064–1069.

Hauser, Marc D. (1996). *The Evolution of Communication*. MIT Press.

Heath, Chip (1995). Escalation and De-Escalation of Commitment in Response to Sunk Costs: The Role of Budgeting in Mental Accounting. *Organizational Behavior and Human Decision Processes*, *62*(1), 38–54.

Hershey, John C. and Jonathan Baron (1992). Judgment by Outcomes: When Is It

Justified? *Organizational Behavior and Human Decision Processes*, *53*(1), 89–93.

Hogan, Robert and Stephen R. Briggs (1986). A Socioanalytic Interpretaion of the Public and the Private Selves. In Roy F. Baumeister (Ed.), *Public Self and Private Life* (179–188). Springer-Verlag.

Holton, Richard (2009). *Willing, Wanting, Waiting*. Oxford University Press.

Hutto, Daniel (2007). *Folk Psychological Narratives: The Socio-Cultural Basis of Understanding Reasons*. MIT Press.

Kahneman, Daniel and Amos Tversky (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, *47*(2), 263–292.

Kausel, Edgar E., Santiago Ventura, and Arturo Rodríguez (2018). Outcome Bias in Subjective Ratings of Performance: Evidence from the (Football) Field. *Journal of Economic Psychology*. Advance online publication. https://doi.org/10.1016/j.joep.2018.12.006

Kelly, Tom (2004). Sunk Costs, Rationality, and Acting for the Sake of the Past. *Noûs*, *38*(1), 60–85.

Konig-Kersting, Christian, Monique Pollmann, Jan Potters, and Stefan T. Trautmann. Good Decisions vs. Good Results: Outcome Bias in the Evaluation of Financial Agents. Working paper, Tilburg University. Retrieved from https://www.uni-heidelberg.de/md/awi/professuren/behfin/kppt_outcomebias_21dec2017.pdf

Korsgaard, Christine (2009). *Self-Constitution*. Oxford University Press.

Kurzban, Robert (2010). *Why Everyone (Else) Is a Hypocrite: Evolution and the Modular Mind*. Princeton University Press.

Kurzban, Robert and C. Athena Aktipis (2007). Modularity and the Social Mind: Are Psychologists Too Self-ish? *Personality and Social Psychology Review*, *11*(2), 131–149.

Langer, Ellen J. (1975). The Illusion of Control. *Journal of Personality and Social Psychology*, *32*(2), 311–328.

Leary, Mark R. (2007). Motivational and Emotional Aspects of the Self. *Annual Review of Psychology*, *58*, 317–344.

Lefgren, Lars J., Brennan Platt, and Joseph Price (2014). Sticking with What (Barely) Worked: A Test of Outcome Bias. *Management Science*, *61*(5), 1121–1136.

Levine, Sheen S. and Robert Kurzban (2006). Explaining Clustering Within and Between Organizations: Towards and Evolutionary Theory of Cascading Benefits. *Managerial and Decision Economics*, *27*(2/3), 173–187.

Loewenstein, George and David Adler (1995). A Bias in the Prediction of Tastes. *Economic Journal*, *105*(431), 929–937.

Mankiw, Gregory (2004). *Principles of Microeconomics* (3rd ed.). Thomson South-Western.

McAfee, R. Preston, Hugo M. Mialon, and Sue H. Mialon (2010). Do Sunk Costs

Matter? *Economic Inquiry*, *48*(2), 323–336.

McClennen, Edward (1990). *Rationality and Dynamic Choice: Foundational Explorations*. Cambridge University Press.

McGeer, Victoria (2007). The Regulative Dimension of Folk Psychology. In Daniel D. Hutto and Matthew M. Ratcliffe (Eds.), *Folk Psychology Re-Assessed* (137–156). Springer.

McKenzie, Richard and Dwight R. Lee (2006). *Microeconomics for MBAs: The Economic Way of Thinking for Managers* (2nd ed.). Cambridge University Press.

Moon, Henry (2001). Looking Forward and Looking Back: Integrating Completion and Sunk-Cost Effects Within an Escalation-Of-Commitment Progress Decision. *Journal of Applied Psychology*, *86*(1), 104–113.

Nozick, Robert (1993). *The Nature of Rationality*. Princeton University Press.

Parfit, Derek (2001). Rationality and Reasons. In Dan Egonsson and Ingmar Persson (Eds.), *Exploring Practical Philosophy: From Action to Values* (17–39). Ashgate.

Parfit, Derek (2011). *On What Matters*. Oxford University Press.

Pettit, Philip (1991). Decision Theory and Folk Psychology. In Michael Bacharach and Susan Hurley (Eds.), *Foundations of Decision Theory* (147–175). Basil Blackwell.

Phillips, Owen, Raymond C. Battalio, and Carl A. Kogut (1991). Sunk and Opportunity Costs in Valuation and Bidding. *Southern Economic Journal*, *58*(1), 112–128.

Rosenberg, Morris (1979). *Conceiving the Self*. Basic.

Ross, Don (2005). *Economic Theory and Cognitive Science*. MIT Press.

Schlenker, Barry R. (1975). Self-Presentation: Managing the Impression of Consistency when Reality Interferes with Self-Enhancement. *Journal of Personality and Social Psychology*, *32*(6), 1030–1037.

Schlenker, Barry R. (1980). *Impression Management: The Self-Concept, Social Identity, and Interpersonal Relationships*. Brooks Cole.

Schlenker, Barry R. (Ed.) (1985). *The Self and Social Life*. McGraw-Hill.

Schoorman, F. David, Roger C. Mayer, Christina A. Douglas, and Christopher T. Hetrick (1994). Escalation of Commitment and the Framing Effect: An Empirical Investigation. *Journal of Applied Social Psychology*, *24*(6), 509–528.

Schrauger, J. Sidney and Thomas J. Schoeneman (1979). Symbolic Interactionist View of Self-Concept: Through the Looking Glass Darkly. *Psychological Bulletin*, *86*(3), 549–573.

Staw, Barry M. and Frederick V. Fox (1977). Escalation: The Determinants of Commitment to a Chosen Course of Action. *Human Relations*, *30*(5), 431–450.

Staw, Barry M. (1976). Knee-Deep in the Big Muddy: A Study of Escalating Commitment to a Chosen Course of Action. *Organizational Behavior and Human Decision Processes*, *16*(1), 27–44.

Staw, Barry M. and Ha Hoang (1995). Sunk Cost in the NBA: Why Draft Order Affects Playing Time and Survival in Professional Basketball. *Administrative Science Quarterly*, *40*(3), 474–494.

Steele, David R. (1996). Nozick on Sunk Costs. *Ethics*, *106*(3), 605–620.

Stone, Jeff, Andrew W. Weigand, Joel Cooper, and Elliot Aaronson (1997). When Exemplification Fails: Hypocrisy and the Motive for Self-Integrity. *Journal of Personality and Social Psychology*, *72*(1), 54–65.

Swann, William B. (1985). The Self as Architect of Social Reality. In Barry R. Schlenker (Ed.), *The Self and Social Life* (100–125). McGraw-Hill.

Tan, Hun-Tong and J. Frank Yates (1995). Sunk Cost Effects: The Influence of Instruction and Future Return Estimates. *Organizational Behavior and Human Decision Processes*, *63*(3), 311–319.

Tarvis, Carol and Elliot Aaronson (2007). *Mistakes Were Made (But Not by Me): Why We Justify Foolish Beliefs, Bad Decisions, and Hurtful Acts*. Harcourt.

Tedeschi, James T., Barry R. Schlenker, and Thomas V. Bonoma (1971). Cognitive Dissonance: Private Ratiocination or Public Spectacle? *American Psychologist*, *26*(8), 685–695.

Thaler, Richard (1980). Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior and Organization*, *1*(1), 39–60.

Tice, Dianne M. and Roy F. Baumeister (2001). The Primacy of the Interpersonal Self. In Marilynn B. Brewer and Constantine Sedikides (Eds.), *Individual Self, Relational Self, Collective Self* (71–88). Psychology.

Tooby, John and Leda Cosmides (1996). Friendship and the Banker's Paradox: Other Pathways to the Evolution of Adaptation for Altruism. *Proceedings of the Royal British Academy*, *88*, 119–143.

Trivers, Robert L. (2000). The Elements of a Scientific Theory of Self-Deception. *Annals of the New York Academy of Science*, *907*(1), 114–131.

Velleman, J. David (2005). The Self as Narrator. In John Christman and Joel Anderson (Eds.), *Autonomy and the Challenges to Liberalism: New Essays* (56–76). Cambridge University Press.

Velleman, J. David (2009). *How We Get Along*. Cambridge University Press.

Whyte, Glen (1986). Escalating Commitment to a Course of Action: A Reinterpretation. *Academy of Management Review*, *11*(2), 311–321.

Wolfers, Justin. Are Voters Rational? Evidence from Gubernatorial Elections. Working Paper 1730, Stanford Graduate School of Business. Retrieved from https://www.gsb.stanford.edu/faculty-research/working-papers/are-voters-rational-evidence-gubernatorial-elections

Zakay, Dan (1984). The Evaluation of Managerial Decisions' Quality by Managers. *Acta Psychologica*, *56*(1-3), 49–57.