How to Do Things with Sunk Costs

Michael Zhao

(forthcoming in *Noûs*)

Abstract

It is a commonplace in economics that we should disregard sunk costs. The sunk cost effect might be widespread, goes the conventional wisdom, but we would be better off if we could rid ourselves of it. In this paper, I argue against the orthodoxy by showing that the sunk cost effect is often beneficial. Drawing on discussions of related topics in dynamic choice theory, I show that, in a range of cases, being disposed to honor sunk costs allows an agent to mimic a *resolute chooser*, someone who adopts the best plan at the outset of a decision problem and sticks with it, even when resoluteness is unfeasible. I discuss several kinds of cases in which honoring sunk costs coincides with resolute choice.

1 Sunk costs

An aircraft manufacturer has invested billions of dollars and a decade in the development of a supersonic aircraft. It has gradually become apparent to the company that the aircraft is unlikely to be profitable: not only will the company be unable to recoup their past investment, but the project will also likely never turn a profit. Had the company had this information before they actually started the project, they never would have started it. Still, the company takes the fact that it invested significant resources on the project as reason to continue investment in the project: we can't let the money and time already invested on it go to waste, it might think.[1]

I would like to thank Zach Barnett, Harjit Bhogal, Tom Kelly, and two anonymous reviewers for *Noûs* for helpful comments on this paper.

[1] This fictional example is obviously based on the real-life example of the development of the Concorde, undertaken by French and British manufacturers and the governments of the two

The aircraft manufacturer exhibits what is known as the *sunk cost effect*. *Sunk costs* are ones that have already been incurred in the pursuit of some end: the money that I spent on a movie ticket, the years that a student spent on a degree, and the billions that the company spent on the project are all examples of sunk costs. When an agent has incurred a sunk cost, she can often *honor* the sunk cost by continuing to pursue the goal for which the cost was incurred, as the company does by continuing to invest in the project; if some good eventually arises—although not necessarily one that is proportional to the cost incurred—then that cost will "be redeemed," or "not have been for nothing." The sunk cost effect is the disposition to continue along courses of action that one has invested significant resources along *because* of those past investments.[2]

countries. Although the project went massively over budget and schedule, costing about about £2 billion (equivalent to £14 billion in 2023) and taking 14 years, and despite pessimistic signs about the future profitability of the aircraft, the parties continually invested more money into the project. It has often been suggested that their reason for doing so was to prevent the past investments from having been in vain. Dawkins and Carlisle (1976, 131), for example, write, "A government which has invested heavily in, for example, a supersonic airliner, is understandably reluctant to abandon it, even when sober judgment of future prospects suggests that it should do so." But there is little evidence for this, and a likelier explanation has to do with the fact that the countries stood to gain a great deal of prestige by seeing the project through.

[2] The description of the sunk cost effect raises a few questions. First, what exactly is a *course of action*? I think of a course of action as containing all of the actions that an agent performs in carrying out a particular intention. Intentions can have a nested structure, and this means that courses of action can be nested inside one another as well. The aircraft manufacturer might intend ultimately to design a profitable supersonic aircraft, intend as a subsidiary goal to develop a supersonic airplane engine, intend as a subsidiary goal to that goal to build an engine testing facility, and so on; each of these intentions generates a different course of action, although those associated with the more subsidiary intentions will be subsets of those associated with the more ultimate intentions. Two actions belong to the same course of action so long as there is some intention they were performed in the execution of, and to continue along a course of action is to continue to (try to) carry out the intention that defines that course.

Second, what is it for an outcome to *redeem* a previously incurred cost? This is a deceptively difficult question. One thought, endorsed by Kelly (2004), is that an outcome redeems a cost when it is a valuable outcome that the cost causally contributed to, even if its value is smaller than the cost: the billions spent on developing the aircraft are redeemed by the eventual success of the aircraft, even if the aircraft is not worth that many billions. But although this might be sufficient for redemption, it is not necessary. After all, the success of the aircraft also redeems dead ends in its development that led nowhere; although they did not causally contribute to the valuable outcome, they are still redeemed because they were pursued *with the intention* of bringing it about. (*This* condition is not necessary either, since a valuable outcome can redeem a cost even if that outcome was unintended: suppose that the airliner proves to be unsuccessful, but the project turns out to have great applications elsewhere.)

The sunk cost effect is widespread. A law school graduate who has spent three years and hundreds of thousands of dollars on his law degree might feel pressured to enter the legal profession, even if he knows there are other professions that are more rewarding and lucrative; he might feel the need to make his degree "not go to waste." A country that has fought a long, costly, and bloody war that seems unwinnable might nonetheless refuse to withdraw, since doing so would mean that its past sacrifices "will have been in vain." More trivially, someone who has bought a new and (as it turns out) unpalatable snack from the supermarket might feel obligated to eat it nonetheless, in order to justify her purchase.

Although the sunk cost effect is a widespread phenomenon, it is almost universally regarded by economists and psychologists as irrational; the standard advice is that one should ignore sunk costs when deciding what to do. In fact, it is commonplace to label the effect a fallacy or cognitive bias. Consider, as a representative example, this passage from a popular economics textbook by Gregory Mankiw (2020, 271):

> At some point in your life you may have been told, "Don't cry over spilt milk," or "Let bygones be bygones." These adages hold a deep truth about rational decision making. Economists say that a cost is a sunk cost when it has already been committed and cannot be recovered. Because nothing can be done about sunk costs, you should ignore them when making decisions about various aspects of life, including business strategy.

Why should it be irrational to care about sunk costs? One common explanation (stated, for example, in the passage by Mankiw) is that we should ignore sunk costs because they cannot be recouped. But this is not quite right. Even when a particular course of action would allow us to recover a past investment,

One might even wonder whether a *valuable* outcome is the only thing that can redeem a sunk cost. After all, it might seem that we can redeem a sunk cost simply by *making use of it* in a significant way, even if the result is not something we would otherwise want. Suppose I book what I think will be an enjoyable beach vacation in Thailand for September before finding out that September is the rainy season in Thailand, so that my vacation is likely to be unenjoyable. Does going on the vacation—something I wouldn't do if I hadn't spend the money—count as *redeeming* the money I spent? (Or do I only *think* I'm redeeming the costs because I *convinced* myself that I'll enjoy the vacation?) For our purposes, we do not need a precise account of redemption. After all, we have an intuitive sense of what motivates people who are moved by sunk costs in these context, and the notion of redemption is supposed to capture whatever they are aiming at here.

that course of action may be worse than a rival course when both are considered in isolation from our past actions; in that case, the standard advice tells us to pursue the second course, even though the first course would allow us to recover the sunk costs.

Suppose, for example, that you have spent $20 on a ticket to see a movie tonight. After buying the ticket, however, you find out from some friends that the movie is boring, so that you would probably have a better time simply by staying at home. As it happens, your ticket is refundable: you can drive to the theater, return the tickets, and recover the $20 that you spent. But doing so would take an hour of your time, and you value the hour of free time more than $20. Here, you have two options. First, you can drive to the theater and get the $20 back, but at the cost of an hour of your time. Second, you can simply do nothing. Although the first option would allow you to recover a past investment, it would incur a new cost greater than the cost it would allow you to recover, and (goes the standard view) it would be rational just to do nothing.

Rather, what makes the sunk cost effect irrational, according to the standard view, is the idea that our actions can affect only the future. On the conception of rationality that we are working with, instrumental rationality, an agent acts rationally if he chooses the option that is the most conducive toward the satisfaction of his ends. Because, the thought goes, all of our choices would have the same effect on the past (none), we should consider only the *future* consequences of our choices when deciding what to do. And so the fact that we incurred significant costs along some course of action in the past provides no reason to continue along that course, since that fact is irrelevant to any of our ends that could still be satisfied. As Lara Buchak (2014, 181) writes, "Allowing one's current preferences to depend on the plans one made in the past rather than on what one desires now is straightforwardly an instance of not taking the currently available means to one's currently desired ends." This is why "the 'sunk-cost fallacy,' treating what one has given up to arrive at a choice node as relevant to what one ought to choose now, is considered a fallacy."

Recently, some have challenged the conventional wisdom, arguing that it need not be irrational to honor sunk costs. Thomas Kelly (2004), for example, has argued that our actions often *can* alter the past—not physically, of course, but they can *change the significance* of past actions or events. In particular, we can make it so that past actions *were not done in vain*. Given that we often have a desire that past actions not have been done in vain, it is often rational to honor sunk costs. I want to pursue another strategy for challenging the convention wisdom. One takeaway of work done in fields like evolutionary psychology, anthropology, and game theory over the last few decades is that many of our tendencies to depart from recognized standards of rationality in how we think or act are actually beneficial in a wide range of cases. As the evolutionary psychologists Leda Cosmides and John Tooby (1994, 329) write, "Despite widespread claims to the contrary, the human mind is not worse than rational… but may often be better than rational." Even if a particular disposition leads us to act in ways that constitute forms of irrationality, perhaps, on the whole, it is still a good thing that we have that disposition.

In this paper, I want to offer a preliminary argument that this is also true for the sunk cost effect: regardless of whether *honoring* sunk costs is ever rational, the *disposition* to honor them can often be beneficial.[3] One reason for this, I will show, is that being so disposed allows the agent to form and execute plans over time that an agent who does not honor sunk costs could not.

2 Dynamic choice

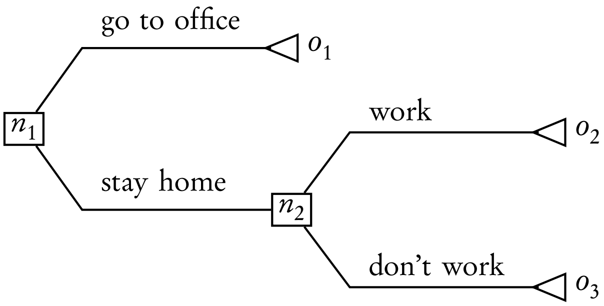
I want to take as my point of departure Robert Nozick (1993)'s brief discussion of sunk costs. Nozick similarly argues that being disposed to honor sunk costs can be beneficial even if actually honoring them is never rational. One key benefit of such a disposition, according to Nozick, is that it often allows us to overcome temptation. To take one of his examples, suppose that I think that seeing many shows at the theatre this year would be good for me; but I also know that, on each particular evening, the temptation of spending a relaxing night at home is always too strong, and I can never bring myself to go to a show. What do I do? Well, I can buy non-refundable tickets to a year's worth of shows in advance. If I am vulnerable to the sunk cost effect, then on each night when there is a show, the thought that the ticket will be wasted if I stay home will motivate me to go see it, allowing me to overcome the temptations of a quiet night in and do what is in my greater interests.

[3] Kelly (2004, §3) endorses this second claim as well, although for game-theoretic reasons. Here, one might wonder whether the fact itself that having a disposition is beneficial makes it *rational* to act on that disposition. There is disagreement: on the one hand, some (like Parfit 1984, ch. 1) take rationality to attach primarily to actions rather than dispositions (or rules for action); on this view, acting on a disposition can be irrational even if the disposition is beneficial. On the other hand, others (like Gauthier 1997, McClennen 1997) take rationality to attach primarily to dispositions or rules for action, so that an action is rationalized by its following from a rational disposition or rule. Since we are not interested in the question of whether actually honoring sunk costs is rational, this is not a question that we need to resolve. In what follows, I'll sometimes speak as if honoring sunk costs is irrational; this is simply for convenience, and should not be read as taking a stand on the debate.

There are many examples of this function of the sunk cost effect. I might buy an expensive gym membership as a way to motivate myself to exercise, knowing that, if I do not, I will have spent all that money on nothing. Or I might actually buy a book that I have been meaning to read, rather than borrowing it from the library, since the fact that I spent money on the book will motivate me to read it so that I will not have spent the money in vain.

Although Nozick does not make such a connection, part of what he has shown is that being susceptible to the sunk cost effect often allows an agent to mimic a principle of choice over time that has been thought to be unfeasible in many cases, for many agents. To spell this point out in more detail, we need to talk a bit about *dynamic choice*, the theory of making decisions over time.[4] A *dynamic choice problem* is a situation in which an agent faces a number of decisions across time, whose outcome is determined by the choices that the agent makes (and possibly external factors). Such a problem has several elements: *choice points*, or occasions on which the agent must choose among different possible actions; *options*, or the (mutually exclusive and collectively exhaustive) choices available at each choice point; and possible *outcomes*. We can graphically represent dynamic choice problems in terms of *decision trees*, in which nodes stand for choice points, branches from those nodes stand for options, and triangles at the end of those branches stand for outcomes.

Consider, for example, the following dynamic choice problem. Suppose that right now, you face the decision either to go to the office or to work from home. If you go to the office, you will have a productive day of work, but you will spend an hour commuting to and from the office. If you stay home, you face a second decision either to work or to slack off: to get distracted by housework, to take a nap, or simply to surf the internet aimlessly. We can represent this dynamic choice problem through the following decision tree:



[4] The exposition in this section largely follows Edward McClennen (1990)'s classic discussion.

The elements of a choice problem generate a set of possible *plans*, or series of directions that tell the agent what to do at each choice point that she will arrive at by following those directions. In this problem, there are three plans: [go to office], [stay home, work], and [stay home, don't work]. Finally, in a dynamic choice problem, the agent typically has *preferences* over the possible outcomes. By "preference," I don't mean anything like a full-fledged judgment about the desirability of different outcomes; rather, to say that the agent prefers X to Y is just to say that she would choose X over Y (even if the agent believes that it is rational for her to choose Y over X). These preferences may change over time, and such changes are part of what make dynamic choice problems distinctive.

We also have to introduce the idea of a *de novo* preference: an agent's preferences for a decision, considering that decision in isolation from anything she has done in the past, as if she were simply thrown into that decision rather than arrived there as the result of previous choices. After all, the fact that an agent has made certain choices in the past or has committed herself to a particular course of action might cause her actual preferences—what she would actually choose—to diverge from what she would choose, considering the decision in isolation from the past. *De novo* preferences are meant to capture the latter idea.[5]

So in the problem above, let us suppose that at n_1, the initial choice point, the agent prefers to get work done from home (o_2) the most; she also prefers to get work done from the office (o_1) over getting nothing done from home (o_3). But if she arrives at n_2 (if she stays home), the prospect of slacking off will be too appealing to her, and—assuming that she does not somehow commit herself to working—her preferences change: at n_2, she prefers not working (o_3) over working (o_2).

Given this dynamic choice problem, how should the agent decide which plan to adopt? One method for deciding is known as *myopic* or *naïve* choice: at each

[5] McClennen actually uses the language of *de novo* decisions; what I call a *de novo* preference is simply a preference in a *de novo* decision. Defining a *de novo* decision more precisely than the gloss I have just given requires some nuance. In a *de novo* decision, I ignore all historical context *except to the extent that it affects the outcomes of my current options*, where those outcomes are defined in terms of their forward-looking features. So, for example, in a *de novo* decision between entering the legal profession and going into another profession, I ignore the time and money that I spent in law school, since those things do not change the outcomes of either option; I treat it as identical to the decision between those two professions that I would face if I had somehow gotten my law degree instantly and for free. In contrast, in a *de novo* decision between drinking water and drinking orange juice, I do not ignore the fact that I just brushed my teeth, since that changes the outcome of drinking orange juice from a pleasant experience to a disagreeable one. (Thanks to a reviewer for this suggestion.)

choice point, the agent should adopt, and act according to, the plan that yields the outcome that she prefers the most at that choice point. In our problem, a myopic chooser would adopt the plan [stay home, work] at n_1, since it yields the best outcome (o_2), given her preferences at n_1. When she arrives at n_2, however, she will prefer o_3 over o_2; thus she will abandon her previously adopted plan, now adopting the plan [don't work], which leads her to o_3. It is plausible that the myopic chooser is irrational: she is led, for completely foreseeable reasons, to an outcome that would not have initially chosen over some other accessible outcome. Had she simply reflected, at n_1, on how she would choose at n_2, she could have arrived at an outcome that she prefered much more.

A second method for choosing, which avoids the lack of foresight exhibited by the myopic chooser, is known as *sophisticated choice*. Sophisticated choice instructs us to anticipate the choices that we will make down the line: first, form expectations about how we will choose at each terminal choice point, given our *de novo* preferences at that point; next, given our expectation about how we will choose at those points, form expectations about how we will choose at each penultimate choice point, given our *de novo* preferences at that point; and so on. Finally, at the initial choice point, choose the option that, in conjunction with how we expect we will choose at each later choice point, yields the outcome that we prefer the most initially. The plan to adopt is that consisting of that choice followed by the choices that one envisions oneself making down the line. So, in our problem, a sophisticated chooser would first anticipate her choice at n_2. Given that she knows that, if she stays home, she will prefer not to work over working, she knows that if she arrives at that choice point, she will choose not to work. At n_1, then, her decision is between the plan [go to office] and [stay home, don't work], since the plan [stay home, work] has been revealed not to be feasible. And of the two plans that she could enact, at n_1, she prefers the outcome yielded by the first to that yielded by the second; so she adopts the plan [go to office].

Finally, at least in theory, there is a method for choosing known as *resolute choice*. Resolute choice tells the agent (1) at the outset, to adopt the plan that yields the outcome that she prefers the most at the outset, and (2) at later choice points, to act according to that plan, even if doing so would go against her *de novo* preferences—what she would prefer had she not adopted the plan—at those points. So unlike sophisticated choice, which constrains the agent's initial choice through her beliefs about how she will choose down the line, resolute choice constrains the agent's later choices through her initial adoption of a plan. In our problem, a resolute chooser would adopt the plan [stay home, work] at the outset, since she prefers o_2 most. When she arrives at n_2, even though the temptation of slacking off is strong enough that, in the absence of her commitment, she would prefer not working over working, she follows through with the plan that she has adopted, choosing to work rather than not to work, which leads her to o_2. In other words, her commitment to the plan *changes her preferences* at n_2 so that, given that commitment, she prefers o_2 to o_3.[6]

3 Is resolute choice feasible?

Obviously, resolute choice requires psychological resoluteness to pull off. For this reason, it may not always be psychologically feasible; agents without the willpower to stick with the plan that they have adopted may not be able to choose resolutely. In such cases, sophisticated choice might be the best remaining option. If I know that I will be unlikely to overcome the temptation of slacking off if I stay home, even if I try to commit myself to working, it seems plausible that I should simply not allow myself to be tempted in the first place, and go to the office.

But there is another problem concerning the feasibility of resolute choice. Notice that in the dynamic choice problem that we have been considering, what resolute choice prescribes at n_2 is simply the prudent option: presumably, the long-term benefits of working outweigh those of slacking off, so what resolute choice counsels us to do at that choice point is simply what an agent who chooses in favor of his long-term interests would do anyway. But this need not be the case. In fact, there are other dynamic choice problems in which sticking to the best plan requires the agent to choose *against* his interests at some later choice point.

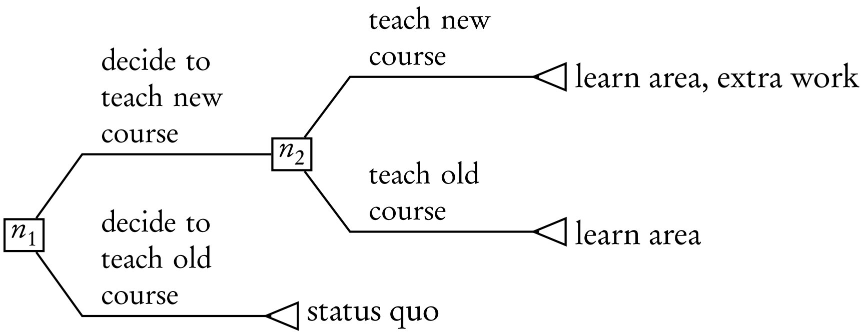
Consider the following problem. Suppose that you have to decide which course you'll teach next semester. Either you can simply decide to teach an introductory philosophy course that you've taught many times in the past, or you can decide to teach a course in an unfamiliar area of philosophy that you've been meaning to learn more about. You don't look forward to the prospect of actually teaching the new course, since it will require a lot of extra work, but the decision to teach the course can serve as the motivation to shore up your knowledge in that area in the meantime.

So suppose that you decide to teach the new course. In the course of preparing for it, you fill the gaps in your knowledge of the subject, think through the

[6] It should be clear from this description that a resolute chooser is someone who not only *intends* to employ resolute choice—the myopic chooser might also intend to do that—but who *succeeds* in employing it.

fundamental questions, and spend a lot of time thinking about how to present the material clearly, thereby deepening your own understanding of it. That is to say, *intending* to teach the course costs a lot in terms of time and energy, but also brings substantial philosophical benefits. As the semester approaches, however, you still have the option to revisit your earlier decision, and to choose to teach intro philosophy. And despite all of the preparation that you have done already, *actually teaching* the new course would still require substantially more effort than simply teaching the old one, without benefiting you enough in compensation. At this point, it would be in your interests to teach the old course rather than the new one.

We can represent this dynamic choice problem through the following tree:



You prefer learning the new area of philosophy without the extra work of teaching the class over learning the new area with the extra work, and you prefer that outcome to the status quo. But the plan that yields the best outcome, [decide to teach new course, teach old course], is not coherent, since it involves intending to do something that you plan not to do. Of the coherent plans, [decide to teach new course, teach new course] is obviously better than [decided to teach old course], so resolute choice tells you to adopt that plan and stick to it. But suppose that you are prudentially rational. The preceding discussion implies that, if you somehow arrive at n_2, you will not actually teach the new course, since doing so is not in your interests. If you are rational, it seems that you cannot act as a resolute chooser would.

Notice that this is a case in which resolute choice is not obviously *psychologically* unfeasible: it does not seem like a case in which choosing resolutely requires tremendous willpower in the face of temptation. Rather, it is one in which resolute choice seems unfeasible given the fact that what it asks the agent to choose down the line is obviously against his interests.

One response is this: Of course, *were* the agent to deliberate again about what to do at the later choice point, he would realize that it makes sense to renege on the plan rather than stick to it. The trick, then, is somehow to get oneself *not* to reconsider one's adoption of the plan down the line. This is the *non-reconsideration* model that has been developed by Michael Bratman (1987, 1999) and Richard Holton (2009). On Holton's model, for example, a resolution consists in both an intention to perform the content of the resolution and a *higher-order* intention not to reconsider the original intention, except under special conditions. Now, Bratman and Holton are concerned primarily with the *rationality* of non-reconsideration rather than with its feasibility. But the claim that non-reconsideration is rational is practically relevant only if it is feasible. And here, Holton writes that we can prevent ourselves from reconsidering the original intention through acts of sheer will: in the face of pressure to reconsider, I exercise willpower and stop myself from doing so, thereby preventing anything that could derail my original intention.[7]

I do not deny that there are agents who follow through with their plans by not reconsidering when they have made up their mind. They might simply have developed a *habit* of not reconsidering once they have formed an intention, for example. And perhaps such a habit could become entrenched to the point that the agent will not reconsider even in the example above, despite how obvious it is that actually following through with the resolution causes gratuitous hassle. My worry is with the broader efficacy of this strategy. If an agent does not already have such a habit, does forming an intention not to reconsider the initial intention actually help him stick to the plan? I do not believe so.

For one, note that it simply pushes the problem a step back. The original problem is that intentions are unstable, given the possibility that I will reconsider them. To preclude this possibility, Holton has me adopt a second-order intention not to reconsider my initial intention. But, of course, this raises the question of the stability of *that* intention, given that I can reconsider *it* too: if I can reconsider my original intention to teach the new course, then I can also reconsider my intention not to reconsider that intention. (Here, it is worth recalling Sartre's idea of the mind as a "nothingness" that can always negate its previous activities; whatever someone resolves, that resolution can be wiped out by an arbitrary act of freedom.)

[7] Holton (2009, 121) writes, "And my suggestion here is that one achieves this [refusal to revise one's resolutions] primarily by refusing to reconsider one's resolutions. On this picture, then, the effort involved in employing willpower is the effort involved in refusing to reconsider one's resolutions."

Second, the non-reconsideration strategy seems better suited to cases of garden-variety temptation, like the decision of whether to work or slack off if I stay home. After all, there are *reasons* not to reconsider my intention in temptation cases that do not apply in the example above, awareness of which bolsters my intention not to reconsider. For example, the mere fact that I am experiencing temptation suggests that, were I to reconsider, my deliberation would be distorted by the temptation; in contrast, my original intention was formed in a cooler hour, when my deliberation was likely much more sound. This is not true in the example above, where I know that my deliberation at the outset is unlikely to be more reliable than my deliberation now.

A further disanalogy is that in temptation cases, refusing to reconsider my original intention does not require me to repress any thoughts, which might be difficult or impossible to do; as Holton writes, I can allow reasons to revise the original intention to enter my head, so long as I do not focus unduly on them. In contrast, it seems that if I am prudentially rational, then as soon as I am aware that actually teaching the new course produces no benefit, I will revise my original intention to teach it. Following the resolution to teach the new course is thus much more difficult than following my resolution (for example) to work from home, since the former requires certain fairly obvious thoughts simply *not to enter my head*.

Finally, there is a body of empirical evidence that shows that people whom we ordinary think of as good at sticking to their resolutions simply have effective habits or employ precommitment rather than actively exercise willpower (Fujita 2011, Galla and Duckworth 2015, Duckworth et al. 2016). It is very easy to stick to my resolution not to snack after dinner if I don't keep snacks at home; it is much harder to do so if I know that the snacks are in the pantry, and I have to exercise willpower to keep myself from eating them. One study, which examined how well students met goals for the semester that they had set for themselves, found that the students who met their goals mostly turned out not to have experienced much temptation in the first place (Milyavskaya and Inzlicht 2017). In fact, the study showed no correlation between self-reports of how much effort students spent on resisting temptation and how successful they were at meeting their goals. The upshot of this is that, if resolute choice—adopting a plan and sticking to it because of one's adoption of it—is supposed to be implemented through an effortful refusal to reconsider the plan that one has adopted, we should have doubts about how generally feasible it actually is.

Again, the conclusion of this section is not that resolute choice is impossible for every agent and in every situation; there are agents who find it easy to stick to their commitments in most situations, and there are situations in which it is easy for most agents to stick to their commitments. Rather, it is simply that resolute choice is unfeasible *for most agents in a wide range of cases*. The unfeasibility of resolute choice in these cases is unfortunate. Consider the teaching example again. Suppose that you know that, even if you somehow decide to teach the new course, you will still just teach the old course at the end of the day. Then you cannot enact the plan [decide to teach new course, teach new course], and the only remaining plan is [decide to teach old course], which yields the status quo. If only you could adopt the best plan, you would receive a large benefit at a small cost; but if you know that you cannot choose resolutely at the later choice point, then you cannot adopt that plan, and you deprive yourself of the opportunity for that benefit.

Paradoxically, a myopic chooser would do the best of all in this problem: he would begin by adopting the plan [decide to teach new course, teach old course], which would result in his learning the material that he wanted to learn; at n_2, however, he would switch to the plan [teach old course], which would save him from all of the hassle of actually teaching the new course. A lack of foresight would actually be a blessing in a such a case. But given that most of us *do* possess the foresight to anticipate our later choices in cases like this, the unfeasibility of being resolute seems to mean that we are condemned to being sophisticated.

4 Redemptive choice and resolute choice

What is the relevance of the sunk cost effect for all of this? My claim is that being susceptible to the sunk cost effect can, in a range of cases, allow the agent to *mimic* a resolute chooser, following through with optimal plans when doing so would otherwise be unfeasible. This is true because resolute choice and honoring sunk costs both violate a principle known as *separability* (McClennen 1990): that the agent's *de novo* preferences, what she would choose when considering the decision in isolation from anything she has done in the past, should dictate her actual preferences. When faced with a decision, separability enjoins us to ignore the historical context that led to that decision; it says to think about what we would choose if we were simply thrown into that decision with fresh eyes.

A sophisticated chooser, given how we have defined her, respects separability; she makes her choices only on the basis of her *de novo* preferences at each point and her expectations of her future choices. But in the case of resolute choice, separability is violated by the agent's commitment to a plan, which causes her preferences to diverge from what they would be in the absence of such a commitment. And in the case of someone who honors sunk costs, separability is violated by the fact that the agent's previous investments motivate her to honor those investments, which similarly creates a divergence between her actual preferences and those she would have in the absence of her past investments. In the case of both, the violation of separability gives the agent a kind of momentum, allowing her to stay on some course of action that she has already begun, despite the fact that doing so conflicts with her *de novo* preferences in some decisions. And *knowing* that she has the ability to choose contrary to her *de novo* preferences down the line can allow the agent clearheadedly to adopt certain plans that she otherwise could not.

This does not mean that someone who honors sunk costs will *always* choose as the resolute chooser would. And there are other cases where a resolute chooser would end up at the same outcome as someone who honors sunk costs *without* having to incur a cost along the way. It also matters how much sunk costs matter to the agent: an agent who merely feels a tinge of regret at having past actions be in vain may be indistinguishable in his choices from an agent who does not care about sunk costs at all. So I'll define *redemptive choice* as another method of choice, meant to capture (with some idealization) the deliberation of an agent to whom sunk costs matter a great deal.[8] Like sophisticated choice, redemptive choice has the agent anticipate his future choices and make choices in the current decision so that, in conjunction with those future choices, he ends up enacting the best plan; to that extent, it constrains his current choices with his beliefs about his future choices. Unlike sophisticated choice, however, redemptive choice does not require that an agent's preferences at a decision be his *de novo* preferences, what he would prefer in isolation from his previous choices. Rather, redemptive choice prescribes the option that allows the agent to redeem a sunk cost, when such an option is available. In this way, redemptive choice also constrains an agent's future choices with his current choices.[9]

[8] This notion has much in common with Kelly's notion of a *pure honorer* of sunk costs, someone whose motivation to redeem a sunk cost varies proportionally to how large that cost was (and is typically significant).

[9] Like the other methods of dynamic choice, redemptive choice is not defined for cases where the agent *gains information that changes the decision problem* after the outset (for example, by revealing that one of the outcomes is not as good as the agent originally thought), which is where the sunk cost effect often manifests. This raises of question of how to extend redemptive choice to such cases. If a redemptive chooser learns that the movie is no good after already buying tickets for it, will he see the movie nonetheless?

It's not terribly important how we answer this question, since the cases that we are concerned with are ones in which the decision problem does not change. Nonetheless, given that redemptive

More precisely, we can define redemptive choice in terms of the following rules, which are meant to be applied starting from terminal choice points and moving backward:

1. At each choice point, if one of the options allows the agent to redeem a cost incurred by a previous choice, the agent should choose that option.

2. Otherwise, the agent should choose the option that, in conjunction with the options that he will choose at any later choice points, leads to the outcome that he prefers *de novo* the most at that point, assuming that his choices at later points follow these two rules.

Obviously, rule 1 can be made more precise a number of ways. What if two options give the agent a chance to redeem a sunk cost? What if they let him redeem different sunk costs? The answers to these questions will not matter for what follows. But I will stipulate, first, that if no option allows the agent to redeem a sunk cost *with certainty*, he should choose the option that gives him a *chance* of redeeming a sunk cost if no other option does so; and second, that if different choices allow an agent to redeem a sunk cost, he should choose the one that *most fully* redeems that cost, where this is a matter of using that cost to the greatest effect.

Rule 2 demonstrates the similarity between redemptive choice and sophisticated choice in their use of backward induction to determine what the agent should choose. Rule 1 is what makes redemptive choice distinctive, in allowing costs incurred by the agent in previous choices to determine what he should do in a given decision. Framed in these terms, the point that I want to make is not that resolute choice and redemptive choice always coincide, but only that they coincide in certain widespread, natural classes of cases. Since they often coincide, redemptive choice can offer the agent of the benefits of resolute choice, even in cases where resolute choice is unfeasible.

In the remainder of this section, I want to look at three broad, natural classes of cases in which redemptive and resolute choice coincide. The prevalence of

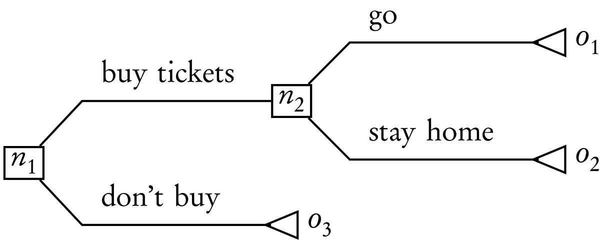
choice is meant to capture how ordinary people who care about sunk costs deliberate, it is in the spirit of the proposal to suggest that the redemptive chooser will seek to redeem the sunk cost even if he learns that the redemptive choice is less valuable than he originally thought. Extending redemptive choice in this way also makes it more robust. After all, in the real world, we can often expect that we'll gain *some* information about the value of different choices. The redemptive chooser can clearheadedly precommit himself to a choice down the line, given the possibility of receiving such information, only if he will choose the option that redeems a sunk cost even in the light of such information.

these kinds of cases establishes a substantial benefit to being susceptible to sunk costs.

Long-term projects

First, being susceptible to sunk costs can allow agents to follow through with long-term projects, and to adopt these projects in the first place. I want to build up to these cases by first recalling Nozick's theatre case; although not an example of a long-term project, it shares certain similarities with those cases that will help us understand how redemptive choice mimics resolute choice there.

So recall the theatre case: I think that it will be good for me to see theatre shows, but I know that, on each evening, I may give in to the temptation of staying home. The problem can be represented through the following decision tree:



Consider the deliberation of a sophisticated chooser, who is not susceptible to the sunk cost effect. Right now, he prefers o_1 (seeing the shows) most of all, and prefers o_3 (staying at home without having spent money) over o_2 (staying at home, but having spent money on tickets); in a *de novo* decision at n_2, however—that is, if he somehow finds himself with tickets—then he will prefer o_2 (staying at home) over o_1 (seeing the shows). If he is a sophisticated chooser, then the knowledge that, at n_2, his *de novo* preference is to stay home means that, at n_1, he will choose not to buy the tickets in the first place. If he were a resolute chooser, he would adopt the plan [buy tickets, go], and when the evening of the show approaches, he would see the show in order to follow through with the plan. But as we noted, this requires quite a bit of willpower on the agent's part, which may not be available to most people.

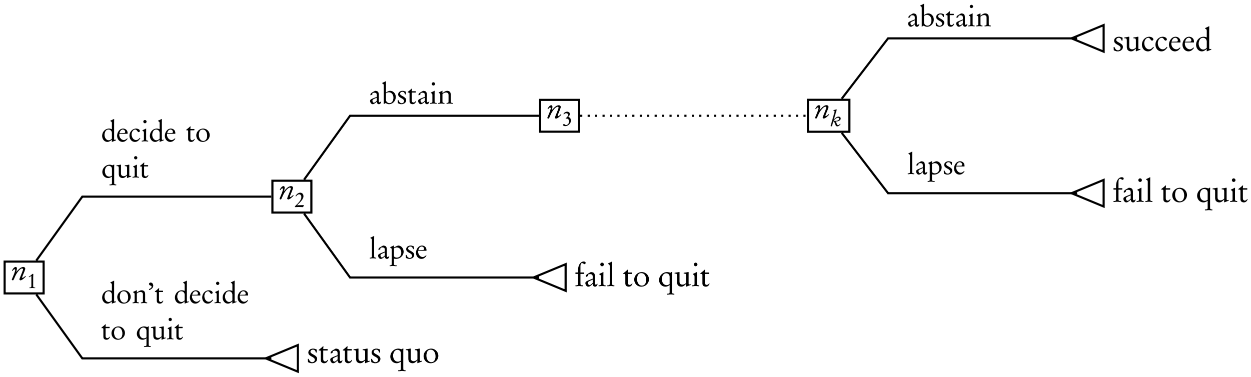
On the other hand, consider a redemptive chooser: one who both has foresight and honors sunk costs. At n_2, he knows that going to the show will redeem his earlier purchase of the tickets. Even though in a standalone decision, he would prefer to stay home rather than go, his previous investment alters his preferences so that, given that investment, he prefers to go rather than to stay home. At n_1, he knows that he will make this choice at n_2, so he chooses to buy the tickets, since that choice, in conjunction with the choice that he knows that he will make later, leads him to the outcome that he prefers the most. In other words, he can take advantage of his susceptibility to the sunk cost effect to *precommit* himself to seeing the show at n_2 by buying the tickets at n_1, ensuring that he will see it without resolving to do so. In a case like this, then, the redemptive chooser makes choices that are identical to those that a resolute chooser would make, even if the former lacks the resoluteness of the latter.[10]

For our purposes, there are two takeaways from this case. First, being susceptible to sunk costs creates a kind of momentum in the same way that resolute choice does, allowing the agent to overcome the temptation of deviating from the best plan. And second, because the agent *knows* that he has the ability not to succumb to temptation down the line, he can undertake projects, in a clear-eyed way, whose success requires resisting temptation.

So consider the class of decisions involving whether or not to undertake some long-term project whose completion benefits the agent, yet every point in which tempts the agent with abandoning it. As one example, consider the project of giving up a vice like drinking. Even if I decide to quit now, I know that I can always go back on my decision later, and lapse into the vice again. We can represent, with a bit of idealization, the problem as follows:

[10] The idea that the redemptive chooser can precommit himself, in this way, to some choice down the line suggests that he can "bootstrap" the motivation to do almost *anything* simply by incurring some (otherwise unnecessary) cost that the act then redeems. And this observation raises the further question of whether the redemptive chooser *should* bootstrap the motivation to perform a desirable act this way, incurring an unnecessary cost with the aim of precommitting himself to a later choice that will redeem that cost? For example, even if someone offers the agent free tickets to the show, should he still unnecessarily buy tickets with the intention of motivating himself to go see the show, which will redeem the cost of the tickets?

My answer here is: yes, as long as the benefit of the redemptive choice minus the cost incurred is larger than the benefit of any alternative option. Perhaps it sounds strange to suggest that someone should incur *unnecessary* costs. But these costs might be unnecessary only in a physical or institutional sense, while being *psychologically* necessary for achieving a good outcome: given features of the agent's psychology, perhaps the only way of achieving that outcome involves incurring those costs. And besides, other precommitment strategies involve incurring costs that are unnecessary in some sense too: strictly speaking, it is unnecessary for me to buy software that blocks internet access while I am working, since I could simply *choose* not to go online then; but it might be a good decision nonetheless. (Thanks to a reviewer for raising this line of questioning.)



Obviously, I prefer successfully quitting to failing. But I also prefer not deciding to quit in the first place to lapsing down the line, since the latter means that I will have expended a lot of effort for nothing long-lasting. Now, I know that I will experience a great deal of temptation to lapse at many choice points in the future; if I were merely dropped into those points, I would almost certainly lapse at some of them. If I am a sophisticated chooser, then given these *de novo* preferences, I expect that I will lapse somewhere down the line, so I simply do not decide to quit in the first place. If I am a resolute chooser, I will decide to quit, and moreover, will stick to this decision at each successive choice point; I thus end up at the optimal outcome. Of course, this might require a tremendous amount of willpower: I have to stick it through just because I decided that I would. But this outcome is also available to a redemptive chooser, one who seeks to redeem sunk costs: I know that at each choice point, I will choose to abstain, since if I lapse, I know that *all of my previous abstentions will be in vain*. "If I drink this time," I might think, "then all of my previous effort—all of those miserable nights out with friends when I stuck with soda—will have been for nothing. So I'd better not ruin all of my previous effort by drinking now." Obviously, for most ordinary agents, this motivation is defeasible; but the point is that someone who honors sunk costs has psychological resources at his disposal that the sophisticated chooser does not, even if he does not have the sheer willpower that the resolute chooser employs.

Here is another example. Suppose that I am deciding whether or not to begin a project, like writing a paper. I prefer finishing the project over not starting it, but I also know that, down the line, I will face frustration at my slow pace or boredom with the tedious parts of the project; if I were simply thrown into those later points with no historical background, I would likely abandon the project. A sophisticated chooser, knowing that he has these *de novo* preferences, will simply never start the project: if I will choose according to my *de novo* preferences, and I will prefer abandoning the project in some later *de novo* decision, then I expect that I will abandon the project somewhere down the line. Given this, it is better not to waste the time and effort in the first place. A resolute chooser will plan to finish the project and, having made that plan, will commit himself to it; his commitment to that plan will get him over the frustration and boredom that some parts of the project will provoke. Finally, a redemptive chooser knows that, even if he feels tempted to give up the project down the line, he will choose not to, since only through finishing the project can he redeem the time and energy he has spent on it. Because he expects that he will choose to continue the project at each point if he begins it, he decides to undertake the project. Again, the sunk cost effect can confer to agents a semblance of resoluteness, allowing the agent to begin and follow through with long-term undertakings when he otherwise would not be able to.

Now, note that in these cases, the choices that resolute choice prescribes are, plausibly, the rational ones: presumably it is in the agent's overall interests not to lapse or to continue working on the project. The only reason that an agent might not choose these things is because of weakness of will, a paradigmatic form of irrationality.[11] As Nozick puts it, even if honoring sunk costs is irrational, in these cases, such irrationality can be exploited to combat another form of irrationality, so that the agent ends up making the rational choice. But can a disposition to honor sunk costs be beneficial even when there is no irrationality of this kind? I believe it can. The next two classes of cases are ones in which redemptive choice and resolute choice coincide, even though at some of the later choice points, it is not irrational (and perhaps even rationally required) to make a choice that violates those methods of choice.

Risky choices

A second kind of case in which the sunk cost effect can be beneficial, by allowing the agent to mimic resolute choice, is one in which *risk attitudes* come into play. It is well known that people are typically *risk-averse*: that is, they prefer getting a payoff for certain to a gamble that, on expectation, produces the same payoff. For example, most people would prefer to win $100 for certain to having a 50%

[11] Though cf. Holton (2009), who argues that giving in to temptation may be rational. After all, when confronted with temptation, the agent's desires and other subjective motivations might shift to a degree that what would best satisfy them is doing what one is tempted to do. Perhaps, when craving a glass of whiskey, the agent who previously wanted to quit drinking cares about nothing more than having a drink to a degree that it swamps all other concerns.

chance of winning $200 and a 50% chance of winning nothing, even though the expected monetary value of that risky outcome is also $100. In fact, most people are risk-averse enough that they would prefer the status quo over certain gambles that could turn out either way, but that, on expectation, result in a positive payoff. For example, most people would prefer not to play a game in which you win $200 if a fair coin lands heads and lose $100 if it lands tails, even though the expected monetary value of doing so is positive ($50).

A *lottery* is the set of possible outcomes of a given choice paired together with their probabilities. So the lottery yielded by playing the game is: gain $200 with probability 1/2, lose $100 with probability 1/2. Now, a well-known fact about risk-averse agents is that even if they would not accept a lottery with positive expected monetary value, they would likely accept a *large number of independent repetitions* of that lottery. As the economist Paul Samuelson (1963) famously discussed, even though most people would reject the lottery above, they would accept 100 repetitions of it offered as a single package. The reason for this has to do with the law of large numbers, which implies that it is very likely that the average outcome of a large number of independent repetitions of some lottery is close to its expected value; in essence, the riskiness of the average outcome decreases with each repetition. After 100 repetitions of Samuelson's lottery, for example, it is very likely that you will have won close to $5,000, and there's less than a 3% chance that you will have lost money. Most people are comfortable with that degree of risk.

There has been a lot of discussion of risk aversion in decision theory and economics, and most don't think that being risk-averse is a kind of *irrationality*, although some neighboring phenomena are irrational. But risk aversion does lead to a potential conflict between one's preferences for the *single case* and one's preferences for the *long run*: a risk-averse agent has *de novo* preferences for single-case situations that might lead him to make a series of choices that, taken as a whole, constitute making a choice over the long run that he disprefers. A risk-averse agent might reject Samuelson's lottery on each occasion that it is offered, even if he knows that it will be offered 100 times; but if he does this all 100 times, then he will have rejected a compound lottery that he would have accepted had it been offered to him all at once.[12]

Similarly, consider a lottery that, for some upfront cost, offers a low chance of a wonderful outcome. Suppose that a cashier in a supermarket is deciding

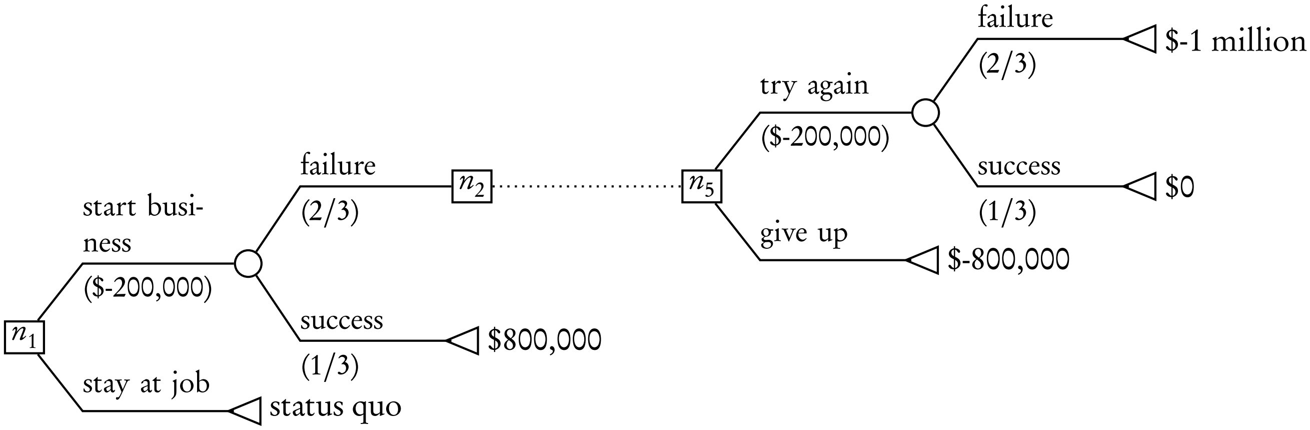
[12] In other words, the agent might violate a principle that McClennen (1990) calls *normal form–extensive form convergence*.

whether to stay in her current job (steady but uninteresting, with limited opportunities for advancement) or to quit in order to start her own grocery store. Of course, choosing the latter is quite risky: the chance that the business will succeed is fairly low, say, one in three, and starting it entails putting up a large chunk of capital up-front and forgoing the income that she would earn otherwise. Suppose that the costs (including the opportunity costs) of starting the store amount to $200,000, while the profit that one would earn if the business succeeds is about $1 million. The lottery yielded by leaving her job to start her own business has positive expected monetary value, but it is risky enough that, in a standalone decision, a risk-averse agent might not choose it.

On the other hand, suppose that the agent has enough savings and possible sources of capital (friends, family, small business loans) to fail up to five times. If the business fails, she can try again: perhaps she can close the old store and buy a store in a different location, or she can try starting another kind of business. At any point, if she grows discouraged, she can always return to her old job, or something similar. Now the situation might seem quite different. After all, if she tries as many times as possible, then it is very likely (there's a 87% chance) that one business will be successful, in which case she will do at least as well as if she had stayed at her job. Of course, choosing to try as many times as possible is still risky; but given that it offers a high chance of success, even many moderately risky-averse agents would be willing to accept the lottery that it yields.

The problem, however, is that the agent confronts the choice to try as many times as possible not as a *single* choice, but as a *series* of choices: on each occasion, she can choose whether to try again, or whether to stop trying and return to his old job. There is no guarantee, then, that she will end up making the choices that she would make were she to face all of the decisions as a single package.

We can represent the problem as follows, where the circular nodes represent "decisions" made by chance:



What would a sophisticated, moderately risk-averse agent do? If she has already started four businesses that have failed, then she knows that she has enough capital only to start one more business. Given that the one-time lottery is quite risky, she has a *de novo* preference not to start it; and given that her *de novo* preferences dictate her actual preferences, she will not start it. But since she expects that at n_5, she will not start another business, then if she arrives at n_4, she knows that *that* will be the last business that she might start. And again, given how risky choosing to start it is, she chooses not to. Continuing this reasoning, at n_1, she expects that she will not start any other businesses, so that the first one is the only one she might start. And again, given how risky choosing to start it is, she chooses not to. This is unfortunate, since, considering the problem as a whole, the agent prefers that she start as many businesses as possible. But in reasoning in a sophisticated way, she ends up adopting a different plan from that.

On the other hand, a resolute agent would simply adopt the plan to start as many businesses as possible, since that plan expresses her preference at the outset.[13] Of course, if she reaches the final choice point, following that plan violates the preferences that she would have in the absence of that commitment. And one might think that, for this reason, most people could not act resolutely here. After all, imagine how someone who doesn't honor sunk costs might reason if she arrives at n_5: "I've already started four businesses, each of which failed, and I only have enough money left to try one more time. If I start another business, it will probably fail, and it's not worth the money that I have to put in and the income that I'm forgoing by choosing for a less-than-even chance of success. I know that I decided years ago that I would try as many times as possible, but

[13] See Thoma (2019) for a more detailed discussion of how resolute choice allows risk-averse agents to align their single-case preferences to their long-run ones.

why should that matter now? I might as well just cut my losses."

Even if resolute choice is unfeasible for many people, though, redemptive choice leads the agent to choose identically in this case. Consider a redemptive chooser: If she arrives at the final choice point, then she knows that all of her past investment—all of the money that she has spent, not to mention the steady income she has forgone—will be in vain if it does not result in a successful business. On the other hand, a successful business will redeem her efforts; her decision to quit her old job will finally have been worth it if she manages to pull it off. It is true that the investments in her previous businesses will not have causally contributed to her success. But the ultimate goal for which those investments were made—to succeed in business—will have been realized, which will redeem them as well. The chance that she will succeed on this last try is small, but it is not nothing; and, at any rate, starting another business will put all of the acumen she has developed over the years to good use. This leads her to try again, despite her risk aversion, since trying again gives her the only chance that she has to redeem her past effort. Similarly, at n_4, n_3, and n_2, she will try again to have a chance of redeeming her earlier efforts. At the initial choice point, knowing that she will continue trying despite her risk aversion, she chooses to leave her job and start a business. In this way, the honorer of sunk costs enacts a plan that is much better, considered from the viewpoint of the problem as a whole, than the plan that the sophisticated chooser enacts.

Of course, this case is one instance of a larger class of cases: those in which an agent faces a repeated decision between the status quo and a lottery that has positive expected value, which imposes a cost in exchange for a small chance of a wonderful outcome that will redeem that cost (and costs incurred by previous instances of that choice). And such cases are not uncommon in real life. Consider the decision about whether or not to apply to a particular fellowship: it is great if you win it, but the application process is quite onerous, and the chance of success is very small. Perhaps the expected value of applying is still positive, but many people are risk-averse enough that the very small chance of winning the fellowship would put them off from applying. Even if you do not win that particular fellowship, however, there are still others that you can apply to. And suppose that, were you to apply to all of them, it is very likely that you would win one. If that is so, then even a moderately risk-averse agent would prefer to apply to as many as possible, when thinking about the problem as a whole. But given that he does not face all of the decisions at once, but rather in sequence, how can the agent be sure that he actually will apply as many times as possible?

Again, being a redemptive chooser—that is, giving sunk costs serious weight—

is helpful here. Even if the redemptive chooser has already failed $n-1$ times, the fact that his previous efforts will be in vain if he does not succeed causes him to try the nth time. And similarly, he will try the $n-1$th time even if he has failed $n-2$ times, and so on. At the outset, knowing that he will continue to try in the face of failure, he decides to try to first time, since that choice together with the choices that he knows he will make later constitutes what he regards as the best long-term plan for the decision problem.

Of course, one can imagine cases in which this tendency goes terribly wrong. If the lottery has negative expected value, then even in the long run, it is unlikely that one will be ahead; in those cases, redemptive choice simply leads one to throw good money (or time or energy) after bad. But in cases like the ones above, where one is faced with a series of choices that impose a cost for a small chance of something wonderful, which still have positive expected value, the disposition to honor sunk costs is beneficial by helping to align our preferences for the single case to our preferences for the long run.

Prophatic intention

Finally, I want to consider a class of cases that involve what I will call *prophatic intention*, the phenomenon of intending to do something, ϕ, *as an excuse* to do some other thing, ψ, that is a prerequisite of doing what one intends to do.[14] In such cases, whether one actually ϕs is often irrelevant, since it is ψing that brings the benefits that one seeks; and actually ϕing might be undesirable. Nonetheless, if one foresees that one will not ϕ, then one cannot intend to ϕ in the first place, which means that one cannot derive the benefits of ψing either. These cases are thus a subclass of what have been called *autonomous benefit* cases in the literature, those in which the benefit follows from intending to perform some action rather than from actually performing it. Although the idea of prophatic intention might seem strange, I want to show that it is common, that being capable of intending prophatically is often beneficial, and that being susceptible to sunk costs often allows the agent to intend prophatically, just as a resolute chooser would intend.

Recall the example that we used to illustrate the frequent unfeasibility of resolute choice in §3: For next semester, you can decide to teach intro philosophy again, or you can decide to teach a course in an unfamiliar area of philosophy as an excuse to familiarize yourself with that area in the meantime. As the new term approaches, however, you can revisit your decision; even if you initially

[14] *Próphasis* means excuse in Greek.

decided to teach the new course, you can still switch to teaching intro philosophy. And at this point, even if *intending to teach* the course has motivated you to learn the material, it would still be better for you *actually not to teach* it, since the benefits of intending to teach it have already accrued, and actually teaching it would require more time and energy than the alternative.

Consider a sophisticated chooser. He knows that even if he manages to decide to teach the new course, when he revisits his decision, he will revise it and choose to teach the old course. Knowing that he will not actually teach the new course at the end of the day undermines his intention to teach the new course in the first place, which means that he will simply decide to teach the old course at the outset, depriving himself of the benefits of preparing for the new course. On the other hand, if he is resolute, he will adopt the plan [decide to teach new course, teach new course], and he will follow through with this plan at the later choice point, even though doing so would not be in his interests; together, these choices lead him to a much better outcome. As we mentioned, however, it is unlikely that most agents would be able to follow such a plan: once the benefits have accrued, why take on the extra work for nothing?

Now consider a redemptive chooser. Again, suppose that he manages to decide to teach the new course. After he has finished preparing for the course, he has already invested significant time and energy into it. If he backs out and teaches the old course again, those investments will have partly been in vain. Of course, they won't have been *completely* in vain, given how much he has gained simply from preparing to teach the course. But the choice to teach the new course would *more fully* redeem those investments—it would make the time and energy spent more worth it—than the alternative would. And so at the later choice point, the redemptive chooser will stick to his original decision to teach the new course, even though backing out would serve his interests better. Knowing how he will choose at the later choice point, the redemptive chooser will decide to teach the new course at the outset. His choices result in his gaining a good deal of philosophical knowledge at the cost of actually teaching the course, which he regards as better than the status quo. Again, an agent who honors sunk costs will, in this case, make the same choices that a resolute chooser would.

Note that deciding to teach the course is intending prophatically: the agent *intends* to teach the new course *as an excuse* to do things that offer significant philosophical benefits, even though *actually teaching the course* is irrelevant to those benefits, and is in fact not in the agent's interests. Cases like this abound. I might rent a beach house for later in the summer, even though I dislike going to the beach, since I know that the prospect of being shirtless around strangers will motivate me finally to get in shape. I might plan to present a paper at a conference, even though I dislike conference-going, since the prospect of giving the paper will force me to finish writing it. Or I might decide to walk to the coffee shop and buy a coffee, even though I think the shop's prices are too high, simply to induce myself to take a walk. The urgency or tangibility of what I intend to do acts as a source of motivation, getting me to do something beneficial that I would not otherwise have done.

In each of these cases, I intend to ϕ (go to the beach, present the paper, buy a coffee) as an excuse to perform ψ (get in shape, write the paper, take a walk) that is a prerequisite for ϕing, and that both provides a benefit and incurs a smaller cost that can be redeemed by actually ϕing. Yet once the benefit has accrued, I have no reason actually to ϕ: I could just not go to the beach house once I get in shape, or bail on the conference once the paper is written, or turn around once I reach the coffee shop. Again, a resolute chooser would simply adopt the plan to ϕ at the outset and stick with the plan because of his adoption; but given the challenges that I mentioned, many of us could not choose resolutely in such cases. A sophisticated agent, knowing that he will not follow through with his intention at the later choice point, will simply never form the intention at the outset; in doing so, he deprives himself of the benefit. A redemptive agent—one who honors sunk costs—will follow through with his intention at the later choice point, not because of his earlier commitment, but in order to redeem the cost that he has incurred. Even though the agent does not really want to go to the beach, attend the conference, or buy a coffee, doing so would make tangible use of the effort he has spent. And knowing that he will do so, he can form the intention to ϕ at the outset, allowing himself to derive the benefits of ψing. So in this class of cases, being susceptible to the sunk cost effect is beneficial by allowing us to intend prophatically, just like a resolute chooser.

5 Conclusion

The goal of this paper has been to challenge the received wisdom about sunk costs. While I have not argued that honoring sunk costs is ever rational, I have shown that a disposition to honor them—embodied in the method of choice that I have called *redemptive choice*—is beneficial in three natural and broad classes of cases: ones involving long-term projects, ones involving certain kinds of repeated risky actions, and ones involving prophatic intention. This is because, in these cases, being susceptible to sunk costs allows the agent to mimic a *resolute chooser*, someone who adopts the best long-term plan for choosing and sticks to it, even when doing so is unfeasible. The sunk cost effect can, in this way, act as a surrogate for psychological resoluteness.

This does not settle the question of whether being susceptible to sunk costs is beneficial *all things considered*. After all, we have only looked at the benefits of such a disposition without looking at its costs. Nonetheless, I hope that I have offered preliminary reason to think that the sunk cost effect, like many of our tendencies that lead us to act or think irrationally, might actually be on the whole beneficial, and we might be worse off if we eliminated it from our decision-making.

References

Michael Bratman. *Intention, Plans, and Practical Reasoning*. Harvard University Press, 1987.

Michael Bratman. *Faces of Intention*. Cambridge University Press, 1999.

Lara Buchak. *Risk and Rationality*. Oxford University Press, 2014.

Leda Cosmides and John Tooby. "Better than Rational: Evolutionary Psychology and the Invisible Hand." *The American Economic Review*, 84:327–332, 1994.

Richard Dawkins and T. R. Carlisle. "Parental Investment, Mate Desertion and a Fallacy." *Nature*, 262:131–133, 1976.

Angela Duckworth, Tamar Gendler, and James Gross. "Situational Strategies for Self-Control." *Perspectives on Psychological Science*, 11:35–55, 2016.

Kentaro Fujita. "On Conceptualizing Self-Control as More Than the Effortful Inhibition of Impulses." *Personality and Social Psychology Review*, 15:352–366, 2011.

Brian Galla and Angela Duckworth. "More than Resisting Temptation: Beneficial Habits Mediate the Relationship between Self-Control and Positive Life Outcomes." *Journal of Personality and Social Psychology*, 109:508–525, 2015.

David Gauthier. "Resolute Choice and Rational Deliberation: A Critique and a Defense." *Noûs*, 31:1–25, 1997.

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

Thomas Kelly. "Sunk Costs, Rationality, and Acting for the Sake of the Past." *Noûs*, 38:60–85, 2004.

N. Gregory Mankiw. *Principles of Economics, 9th Edition*. Cengage Learning, 2020.

Edward McClennen. *Rationality and Dynamic Choice*. Cambridge University Press, 1990.

Edward McClennen. "Pragmatic Rationality and Rules." *Philosophy and Public Affairs*, 26:210–258, 1997.

Marina Milyavskaya and Michael Inzlicht. "What's So Great About Self-Control? Examining the Importance of Effortful Self-Control and Temptation in Predicting Real-Life Depletion and Goal Attainment." *Social Psychological and Personality Science*, 8:603–611, 2017.

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

Derek Parfit. *Reasons and Persons*. Oxford University Press, 1984.

Paul Samuelson. "Risk and Uncertainty: A Fallacy of Large Numbers." *Scientia*, 98:108–113, 1963.

Johanna Thoma. "Risk Aversion and the Long Run." *Ethics*, 129:230–253, 2019.