Options and Diachronic Tragedy

Brian Hedden

Philosophy and Phenomenological Research 87 (1)

1 Introduction

In a tragedy, the protagonist suffers some misfortune. What makes this misfortune tragic is that it is foreseeable well before it occurs. In some tragedies, the misfortune is foreseeable only by the audience. But in others the misfortune is in some sense foreseeable by the protagonist himself. The protagonist can foresee that his own desires will drive him to engineer his ruin but nonetheless fails to depart from this disastrous course.

Certain sorts of beliefs and desires, of particular interest to philosophers, are tragic in this second way. They drive you to perform each member of a sequence of actions that you can see will result in a bad outcome, even though there is some alternative sequence of actions that you in some sense could have performed and that would have avoided this bad outcome. In this way, these attitudes lead you to act over time in a manner that is to your own acknowledged, predictable disadvantage. I call this phenomenon *Diachronic Tragedy*.[1]

I begin with a series of cases of Diachronic Tragedy. Each has generated extensive debate, although these debates have been conducted largely independently of one another. In many cases, philosophers have attacked the preferences involved as irrational simply on the grounds that they yield Diachronic Tragedy.

---

[1]Some of these cases have been referred to in the literature as 'money pumps' and 'diachronic dutch books.' I adopt the term 'Diachronic Tragedy' to refer to all cases in which you perform a predictably dispreferred sequence of actions, regardless of whether they involve money (as the term 'money pump' suggests) or betting (as the term 'diachronic dutch book' suggests).

After discussing these cases and highlighting their common structure, I attempt to show that this ubiquitous style of argument - concluding that a certain attitude is irrational from the premise that it yields Diachronic Tragedy - fails. This style of argument crucially relies on the assumption that the rational *ought* can be applied not only to particular actions, but also to sequences thereof. I argue for a theory of a decision-maker's options on which the rational *ought* does not apply to sequences of actions. Therefore, the crucial assumption needed to infer that attitudes which yield Diachronic Tragedy are irrational is false.

I conclude that the mere fact that a certain attitude yields Diachronic Tragedy does not entail its being irrational.[2]

2   Diachronic Tragedy

A *Tragic Sequence* is a sequence of actions S_1 such that at all times you prefer performing some other possible sequence of actions S_2 over performing S_1. Many different kinds of attitudes have the unfortunate consequence that, given those attitudes, you will prefer performing each member of a Tragic Sequence at the time it is available, even though you prefer not to perform the sequence as a whole. Such attitudes are *Tragic Attitudes*. In Sections 2.1-2.7, I present prominent examples of Tragic Attitudes from the literature. I want to emphasize that at this point I remain neutral on the question of whether the attitudes in the examples to follow are in fact irrational. (The reader may skip some of these examples without loss, as the later discussion will focus on their common core, rather than on the specific details of each case.)

2.1   Preference Shifts

Shifts in your preferences can be tragic. Suppose you are the Russian Nobleman imagined by Parfit (1984). You are a 20 year old fervent leftist. But you know that by middle age,

---

[2]In claiming that vulnerability to Diachronic Tragedy does not entail irrationality, I am denying (i) that such vulnerability *constitutes* irrationality and (ii) that such vulnerability is a 100% reliable *symptom* or *indicator* of irrationality.

you will become an equally fervent rightist. Consider:

> **The Russian Nobleman**: You will receive an inheritance of $100,000 at age 60. Right now, you have the option (call it *Donate Early*) of signing a binding contract which will require $50,000 to be donated to left-wing political causes. No matter whether you take this option, you will at age 60 have the option (call it *Donate Late*) of donating $50,000 to right-wing political causes. (No greater donation is permitted under Tsarist campaign finance laws.) Right now, you most prefer donating $50,000 to left-wing causes and nothing to right-wing causes. But you also prefer donating nothing to either side over donating $50,000 to each side, as the effects of those donations would cancel each other out.

Right now, regardless of whether your later self will Donate Late, you prefer to Donate Early.[3] But at age 60, no matter what you do now, you will prefer to Donate Late. But the sequence of actions <Donate Early, Donate Late> is a Tragic Sequence, since at all times, you disprefer it to the sequence <Not Donate Early, Not Donate Late>. It is better to save your money than to give it all away in two donations that cancel each other out.

2.2 Time Bias

Time bias is a special case of predictable preference shifts. You are time biased if you prefer that your pains be in the past and your pleasure in the future, even if this means more pain and less pleasure overall. It was widely thought that time bias was a practically inert sort of preference, since you cannot affect the past. But Dougherty (2011) shows that time bias can make a difference to how you act if you are also risk averse, and that it will do so in a way that leads to tragedy. (On Dougherty's definition, you are risk averse if you prefer a gamble with a smaller difference between the best and worst possible outcomes to a gamble with a higher difference between *its* best and worst possible outcomes, even

---

[3]If your later self does not Donate Late, you would rather give $50,000 to left-wing causes, since this is your most preferred outcome. And if you later self does Donate Late, you would rather cancel the effect of that donation by giving $50,000 to left-wing causes than let that later right-wing donation go unchecked.

if the expected value of the first gamble is somewhat lower than the expected value of the second.)[4]

Suppose you are both time biased and risk averse. Consider:

> **Uncertain Pain**: A coin was flipped to determine which of two surgery regimes you will undergo. If it landed heads, you will have 4 hours of painful surgery on Tuesday and 1 hour of painful surgery on Thursday (the Early Course). If it landed tails, you will have no surgery on Tuesday and 3 hours of painful surgery on Thursday (the Late Course). Either way, you will be given amnesia on Wednesday, so that you won't remember whether you had surgery on Tuesday (though you will remember everything else). There is a clock next to your bed, so you always know what day it is.

On Monday and Wednesday, you will be offered the pills Help Early and Help Late, respectively. Each reduces the difference between the highest possible amount of future pain and the lowest possible amount of future pain:

> **Help Early**: If you are in the Early Course, then taking Help Early will reduce the time of your Thursday surgery by 29 min. If you are in the Late Course, then taking Help Early will increase the time of your Thursday surgery by 31 min.

> **Help Late**: If you are in the Early Course, then taking Help Late will increase the time of your Thursday surgery by 30 min. If you are in the Late Course, then taking Help Late will decrease the time of your Thursday surgery by 30 min.

On Monday, you prefer taking Help Early to refusing it. Why? Because it reduces the difference between the highest and lowest amounts of possible future pain (by reducing the future pain in the Early Course scenario involving the most future pain and increasing the future pain in the Late Course scenario involving the least future pain) at a cost of

---

[4]So, for instance, you might prefer a bet on a coin toss which pays $8.50 if heads and $10.50 if tails to a bet which pays $0 if heads and $20 if tails, even though the expected value of the former bet is $9.50 while the expected value of the latter bet is $10.

increasing your expected future pain by only 1 min. This is true whether or not you take Help Late.

On Wednesday, you prefer taking Help Late. Why? Because it reduces the difference between the highest and lowest amounts of possible future pain without changing your expected future pain at all. This is true whether or not you took Help Early.

But taking both Help Early and Help Late just guarantees you 1 more minute of pain on Thursday than if you had refused both pills. Hence, the sequence of actions <Take Help Early, Take Help Late> is a Tragic Sequence, since at all times you prefer refusing both pills over taking both pills. Hence, time bias is an example of a Tragic Attitude.

2.3   Intransitive Preferences

Suppose you have intransitive preferences. You prefer Apple Pie to Blueberry Pie, Blueberry Pie to Cherry Pie, and Cherry Pie to Apple Pie. Consider:

> **The Money Pump**: You start off with an Apple Pie, a Blueberry Pie, and a Cherry Pie. You will be offered three deals in succession no matter what. Deal 1: receive a Blueberry Pie in exchange for your Cherry Pie and 10 cents. Deal 2: receive an Apple Pie in exchange for a Blueberry Pie and 10 cents. Deal 3: receive a Cherry Pie in exchange for an Apple Pie and 10 cents.[5]

If you act on your preferences at each time, you will be turned into a money pump. You will accept the first deal, giving up 10 cents and a Blueberry Pie in exchange for an Cherry Pie. Why? Because regardless of whether you will go on to accept the second and third deals, you would prefer to move up from Cherry Pie to Blueberry Pie, even at a cost of 10 cents. For perfectly analogous reasons, you will accept the second and third deals as well. But having accepted all three deals, you wind up with the same assortment

---

[5]The original money pump argument is due to Davidson et al. (1955). I have presented an improved version of the standard Money Pump case due to Dougherty (ms). In the standard case, where you start off with Cherry, are then given the opportunity to pay to switch to Blueberry and then again to Apple, and then again back to Cherry. The standard case has the disadvantage of being such that you can avoid ruin by simply refusing the first deal, since the later deal (e.g. paying to switch from Blueberry to Apple) cannot be offered unless you accept the first deal (paying to switch from Cherry to Blueberry). Dougherty's case blocks this escape route, since each deal can be offered no matter whether you accept or decline the deals that were offered before.

of pies that you started with despite your outlay of 30 cents.

The sequence <Accept Deal 1, Accept Deal 2, Accept Deal 3> is a Tragic Sequence, since throughout the whole process, it is dispreferred to the sequence of declining all three deals. So intransitive preferences are an example of Tragic Attitudes.

2.4   Imprecise Preferences

Suppose your preferences are imprecise - You have no preference between a scuba trip to Australia (A) and a safari trip to Botswana (B), but you also do not regard them as equally desirable. For adding \$50 to one of them wouldn't make you then prefer it to the other. You don't prefer $A+$ to B or $B+$ to A (even though you prefer $A+$ to A and $B+$ to B). Imprecise preferences can lead you to misfortune. Consider:

> **Scuba or Safari**: There are two boxes. You see that Box A contains a ticket for the scuba trip A, while Box B contains a ticket for the safari trip B. You know in advance that at t_1 you will get to decide whether \$50 is placed in Box A or Box B, and then at t_2 you will get to take one of the boxes.

You have no preference about which box to put the \$50 in at t_1 (since the situation is symmetric). Suppose you put the \$50 in Box A. Then at t_2 your preferences license you to take either Box A or Box B. In particular, they license you to take Box B (since you do not prefer scuba plus \$50 over the safari). But the sequence <put \$50 in Box A, take Box B> is a Tragic Sequence, since at all times you prefer the outcome (safari plus \$50) that would have resulted from putting the \$50 in Box B and taking Box B. (Similarly, *mutatis mutandis*, for putting the \$50 in Box B and taking Box A.)[6]

---

[6]This argument parallels the argument made by Elga (2010) in the context of imprecise *degrees of belief*. Your degrees of belief are imprecise if there are two propositions A and B such that you do not regard one as more likely than the other, but you also do not regard them as equally likely (so that disjoining one with a proposition with small but positive probability does not make you regard the disjunction as more likely than the remaining proposition). Elga shows that having imprecise degrees of belief can license you to perform each member of a Tragic Sequence.

2.5 Infinite Decisions

One option A *dominates* another option B if and only if option A yields a better outcome than B in every state of the world. It is widely accepted that you are rationally permitted, and even required, to take dominant options. But Arntzenius et al. (2004) argue that in some infinite cases, taking dominant options will lead to trouble:

> **Satan's Apple**: Satan has cut an apple into infinitely many slices. At each of various times t_i, you are asked whether you would like to eat slice $\#i$.[7] If you eat infinitely many slices, you go to Hell, while if you eat only finitely many slices, you go to Heaven. Your first priority is to go to Heaven rather than Hell. Your second priority is to eat as much apple as possible.

For each slice i, eating that slice dominates not eating it. For eating *it* will not make the difference between eating only finitely many slices and eating infinitely many slices, and so it will not affect whether you go to Heaven or to Hell. But if you take the dominant option for each slice - eating it - then you will wind up eating infinitely many slices and be condemned to an eternity in Hell! The sequence of eating every slice is a Tragic Sequence, since it yields a worse outcome than myriad other sequences of actions (e.g., that of refusing every slice). So the preference for dominant options is a Tragic Attitude.[8]

2.6 Formal Epistemology

Formal epistemologists have argued for other principles of rationality on the grounds that violating them can lead you to prefer each member of a Tragic Sequence. I mention three such cases here, without going into full details for reasons of space.

---

[7]The t_i are arranged so that the choosing constitutes a supertask - t_0 is 0 sec from now, t_1 is 30 sec from now, t_2 is 45 sec from now, and so on.

[8]It is an interesting feature of Satan's Apple that each possible sequence of actions is worse than some other sequence of actions. Therefore, even if you had the ability to decide all at once which slices of apple to eat, it is unclear what you ought to do, since whatever sequence you choose, there will be some better one that you could also have chosen. Perhaps there is some threshold number of slices such that you are permitted to choose any sequence in which you eat at least that many slices (but not infinitely many). But any such threshold will inevitably be arbitrary. Perhaps in this case, we must abandon the binary *ought/ought not* distinction in favor of more graded evaluations, in which we can only speak of an action's being more rational than another.

First, Bayesian epistemologists standardly hold that you ought to conditionalize on your evidence. That is, upon learning an evidence proposition E, you ought to set your degree of belief in every other proposition H equal to your prior conditional degree of belief in H given E.[9] Lewis (1999) argues for this principle by showing that if you violate Conditionalization, you will prefer each member of a Tragic Sequence - a clever bookie could offer you different bets at different times, such that each bet looks good at the time it is offered, even though the bets you accept together guarantee you a loss.

Second, van Fraassen (1984) argues for the principle of Reflection, which states that you ought to defer to the beliefs of your later selves. You ought to treat your later selves as experts. You satisfy Reflection if your current degree of belief in H, conditional on your later having degree of belief n in H, is itself n.[10] van Fraassen argues for Reflection by showing that if violate it, you will prefer each member of a Tragic Sequence consisting of accepting bets at different times which together guarantee you a loss.

Third, the Sure-Thing Principle in decision theory requires you to prefer A to B if and only if you prefer a lottery which has A as a possible prize to an otherwise identical lottery with B instead of A as a possible prize.[11] Raiffa (1968) argues that if you violate the Sure-Thing Principle, you will prefer each member of a Tragic Sequence.

2.7 Common Structure

In all of these myriad cases, we can represent your decision situation with this tree:[12]

---

[9]More formally, where P_0 is the probability function representing your initial degrees of belief and P_E is the probability function representing your degrees of belief after becoming certain of E, Conditionalization requires that $P_E(H) = P_0(H|E)$.

[10]More formally, where P_0 is the probability function representing your degrees of belief at t_0 and P_1 is the probability function representing your degrees of belief at a later time t_1, Reflection requires that $P_0(H|P_1(H) = r) = r$.

[11]More formally, the Sure-Thing Principle requires that for all A, B, and C:

$$A \geq B \text{ if and only if } \{A, p; C, 1 - p\} \geq \{B, p; C, 1 - p\}$$

where $\geq$ denotes the *at-least-as-preferred* relation and $\{A, p; C, 1 - p\}$ is a lottery which yields A with non-zero probability p and C with probability $1 - p$.

[12]Of course, the number of decision points required will differ in some cases, like Money Pump (three decision points) and Satan's Apples (infinitely many decision points), but the basic structure is the same.
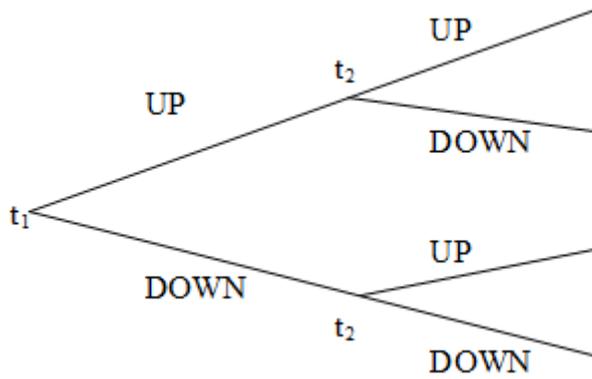
At each of t_1 and t_2, you can either go UP or DOWN. At the first node, you prefer going UP to going DOWN, no matter what you will later do at t_2. That is, you prefer the sequence <UP, UP> over the sequence <DOWN, UP> and prefer the sequence <UP, DOWN> over the sequence <DOWN, DOWN>. And at t_2, you prefer going UP to going DOWN, no matter what you did at t_1. That is, you prefer the sequence <UP, UP> over <UP, DOWN> and prefer the sequence <DOWN, UP> over <DOWN, DOWN>.[13] But at both t_1 and t_2, you prefer the sequence <DOWN, DOWN> over the sequence <UP, UP>. In this way, the sequence <UP, UP> is a Tragic Sequence, but at each time, you prefer performing the member of this Tragic Sequence available at the time.

Let's go through this with a simple example. In **The Russian Nobleman**, Donate Early plays the role of going UP at t_1 and Donate Late plays the role of going UP at t_2. Right now, as a young leftist, you prefer to Donate Early, no matter what you will do later at age 60 (that is, you prefer sequences in which you Donate Early to corresponding sequences in which you don't). But at age 60, you will prefer to Donate Late, no matter what you did as a young leftist. But both now and at age 60, you prefer the sequence <Not Donate Early, Not Donate Late> over the sequence <Donate Early, Donate Late>. Similarly, *mutatis mutandis* for all the other cases in Section 2.

In essence, these cases are Prisoner's Dilemmas, in which your t_1 and t_2 time-slices are the two prisoners. In a Prisoner's Dilemma, prisoners A and B must each choose to defect or cooperate. Each prisoner prefers to defect, no matter what the other will do. That is, prisoner A prefers the 'sequence' of actions <A defects, B defects> over <A cooperates, B

---

[13]The case of imprecise preferences is slightly different. In that case, you do not actually have the preference at each of t_1 and t_2 for going UP; rather, you just lack the contrary preferences at t_1 and t_2. I set this detail aside for the sake of clarity.

defects> and prefers <A defects, B cooperates> over <A cooperates, B cooperates>, and similarly, *mutatis mutandis* for prisoner B. But each prisoner prefers <A cooperates, B cooperates> over <A defects, B defects>. In a Prisoner's Dilemma, then, each prisoner prefers performing one action (defecting) no matter what the other will do but would prefer that neither of them perform that action than that both perform it. In cases of Diachronic Tragedy, each time-slice prefers performing some action but would prefer that neither time-slice perform its preferred action than that both perform it.

Crucially, I have been assuming in all these examples that you lack the ability to *self-bind*.[14] That is, there is nothing you can do *now* which will causally determine which actions your future time-slices will perform. If you have the ability to self-bind and know it, then if you are rational, you will not perform a Tragic Sequence even if you have Tragic Attitudes. This is because, by definition, you prefer performing some other sequence of actions over performing the Tragic Sequence. So, if you know that you can bind yourself to any sequence of actions, then you rationally ought to bind yourself to one of these preferred sequences of actions and thereby avoid performing the Tragic Sequence. (Similarly, if the participants in a Prisoner's Dilemma are able to talk and jointly settle on what they will do, they will each cooperate, since they prefer that they both cooperate rather than both defect.) Because Tragic Attitudes only lead to trouble if you either lack the ability to self-bind or don't know that you have this ability, I will continue to assume in what follows that you are either unable to self-bind or are ignorant of whether you have this ability.[15]

3   The No Way Out Argument

As we have seen, many philosophers have concluded from the fact that a certain attitude (or set of attitudes) yields Diachronic Tragedy that that attitude (or set thereof) is irra-

---

[14]This term comes from Arntzenius et al. (2004).

[15]One might take the Diachronic Tragedy to show not that Tragic Attitudes are irrational, but rather that it is a requirement of rationality that one have the ability to self-bind (and know it); it is a requirement of rationality that one be able to make decisions that bind one's later time-slices to certain courses of action. This line of argument may be supported by the work by Hinchman (2003), Bratman (2007), and Korsgaard (2008), who emphasize the importance to rationality and agency of the capacity to form and execute intentions which guide one's later actions.

tional. This inference is often taken as intuitively compelling and not in need of further argument. But we should be reluctant to simply trust brute intuition here.

First, it is a curious sociological fact that even though in *many* of these cases, philosophers have concluded from Diachronic Tragedy that the relevant attitudes are irrational, in others they have been reluctant to draw the same conclusion. For instance, philosophers have been reluctant to conclude on the basis of Diachronic Tragedy that it is irrational to change your preferences or to have the preferences at issue in **Satan's Apple**.[16] Of course, this sociological fact in no way entails that an inference from Diachronic Tragedy to irrationality is invalid, but it does mean that we should tread lightly and treat arguments based on Diachronic Tragedy with caution.

Second, and more importantly, we already know that certain attitudes can predictably lead to misfortune without being in any way irrational. For example, having an inflated sense of our own talents keeps us happy, motivated, creative, and successful (Taylor and Brown (1988)). But this does not entail that it is irrational to have accurate, evidence-based beliefs about our own talents! (Of course, the fact that a certain sort of attitude is detrimental may entail that it would be rational to desire not to have that attitude, or to attempt to cause oneself not to have that attitude, but this is quite different from entailing that the attitude itself is irrational.) Given these concerns, we must look for a compelling *argument* that Tragic Attitudes are *ipso facto* irrational.

Acknowledging that mere vulnerability to loss does not entail the presence of any irrationality, we might nonetheless think that susceptability to Diachronic Tragedy entails a kind of inconsistency in your attitudes, and that the exploitation involved in cases of Diachronic Tragedy merely serves to make vivid this inconsistency. Compare the dialectic concerning the Synchronic Dutch Book Argument, a close relative of the arguments based on Diachronic Tragedy that are our focus. This argument aims to show that it is irrational to have credences which violate the axioms of the probability calculus on the grounds that such violations will license you to accept each member of some set of bets, even

---

[16]Similarly, Reflection is widely regarded as being subject to numerous counterexamples, which casts doubt on the argument for Reflection based on Diachronic Tragedy. See Christensen (1991) for more on this worry and for discussion of the relationship between the Diachronic Tragedy Arguments for Reflection and conditionalization.

though all the bets taken together logically guarantee you a loss. But arguably this shows only that there are pragmatic disadvantages to violating the axioms, and not that it is epistemically irrational to do so.[17] In response, a number of philosophers have attempted to 'depragmatize' the Synchronic Dutch Book Argument, interpreting it in such a way that it really does show the epistemic irrationality of credences that violate the axioms of the probability calculus. Christensen (1996), for example, argues that credences which violate the axioms 'sanction as fair' each of a set of bets which is together unfair, and that this is an distinctively epistemic problem with the credences themselves. Skyrms (1987) interprets vulnerability to a Dutch Book as showing that you evaluate bets differently depending on how they are described, and that this is likewise a problem with your credences themselves.

I will not seek to evaluate the success of these attempts to depragmatize the Synchronic Dutch Book Argument.[18] Rather, I want to emphasize that this is the route that we should take in attempting to defend the argument that Tragic Attitudes are *ipso facto* irrational; we should seek to show that vulnerability to Diachronic Tragedy means that your attitudes are themselves in some sense inconsistent.[19] Offhand, this is likely to be no easy task. As Christensen (1991) notes, while there is something inconsistent about believing both H and $\neg H$ at one time, there is nothing inconsistent in the diachronic case about believing H at one time and $\neg H$ at another.

But I think that it is not so hopeless to attempt to argue that Diachronic Tragedy illustrates a kind of inconsistency in your attitudes. We might argue that Tragic Attitudes are irrational is that having those attitudes in some cases means that you cannot help but do something you rationally ought not do. And it is in giving rise to conflicting *ought* claims that Tragic Attitudes are in some sense inconsistent.[20]

---

[17]See e.g., Kennedy and Chihara (1979), Rosenkranz (1981), and Joyce (1998).

[18]In Hedden (forthcoming), I argue that depragmatization is beside the point, since the Synchronic Dutch Book Argument is unsound in any event. Credences can violate the axioms of the probability calculus while provable not licensing you to accept each member of a Dutch Book. Such credences involve what I call 'negation incoherence,' in which you credences in a proposition and its negation fail to sum to 1. Negation incoherent credences have been overlooked in the Synchronic Dutch Book Argument, for the argument is based on the assumption that your credences equal your fail betting quotients, and this entails that your credences are negation coherent.

[19]Briggs (2009, 67) for instance talks of 'intrapersonal coherence.'

[20]Again, the case of imprecise preferences is slightly different, in that the relevant preferences seem to

If you have Tragic Attitudes, then either you will perform a particular action that you rationally ought not perform, or you will perform a sequence of actions that you rationally ought not perform. You are caught with no way out. This suggests the following argument that such attitudes are irrational:[21]

The No Way Out Arguman t

P1: A set of attitudes is irrational if there are cases where no matter what you do, you will have done something that, given those attitudes, you rationally ought not have done.

P2: If you have Tragic Attitudes, then in some cases no matter what you do you will have done something that, given those attitudes, you rationally ought not have done.

C: Tragic Attitudes are irrational.

Why believe P1? Rational agents do not do things that they rationally ought not do. So, if having certain attitudes entails that no matter what you do, you will have done something you rationally ought not have done, then you cannot be a rational agent and have those attitudes. A set of attitudes is irrational if you cannot be a rational agent and have those attitudes. Hence, a set of attitudes is irrational if no matter what you do, you will have done something that, given those attitudes, you rationally ought not have done.

Why believe P2? Well, one might argue for P2 by saying that no matter what you do, either you will have performed some particular action you ought not have performed, or you will have performed some *sequence* of actions that you ought not have performed. Consider **The Russian Nobleman**. You rationally ought not perform the sequence of actions <Donate Early, Donate Late>, since at all times you preferred performing some other sequence of actions that was available to you. But you rationally ought to perform the particular action Donate Early, since at the time it is available you prefer to perform

_____

merely *permit*, rather than force, you to do something you rationally ought not do.

[21]While I will later argue on principled, theoretical grounds that this argument is unsound, an initial cause for skepticism is the fact that it does not distinguish among the cases of Tragic Attitudes which do seem irrational (like intransitive preferences) and those which do not (like the preferences in **Satan's Apple**).

it. And you rationally ought to perform the particular action Donate Late, since at the time *it* is available you prefer to perform it. So, you rationally ought to Donate Early, you rationally ought to Donate Late, but you rationally ought *not* perform the sequence <Donate Early, Donate Late>. So you cannot do all that you ought to do. For it is logically impossible for you to Donate Early, Donate Late, but not perform the sequence <Donate Early, Donate Late>. (Similarly for the other cases in Section 2.)

The crucial assumption in support of P2, then, is that the rational *ought* can be applied not just to particular actions like Donate Early and Donate Late, but also to *sequences* of actions like <Donate Early, Donate Late>. But is this assumption correct? I argue that evaluating this assumption, and therefore P2 itself, requires us to get clear about what your options are in a decision situation. We need a general theory of the sorts of things to which the rational *ought* applies. (I use 'option' as a technical term to denote whatever sorts of things are such that you ought or ought not do them in a decision situation, so that it is an analytic truth that if you ought or ought not ϕ, then ϕ-ing is an option for you.)

In the next section, I argue that your options at a time t consist of all and only the decisions you are able to make at t. If this is correct, then sequences of actions are not the sorts of things to which the rational *ought* applies. Hence we cannot say that you rationally ought not perform a Tragic Sequence, since this would involve a category mistake. Moreover, it is possible to have Tragic Attitudes without doing anything that you rationally ought not do. Hence P2 is false, and The No Way Out Argument fails.

I noted in the previous section that cases of Diachronic Tragedy are essentially intrapersonal Prisoner's Dilemmas. Essentially, I will be arguing that we should say about these *intra*personal Prisoner's Dilemmas the same thing that is commonly said about standard *inter*personal Prisoner's Dilemmas. In the interpersonal Prisoner's Dilemma, each prisoner prefers to defect. But by each defecting, they wind up with an outcome which by each of their lights is worse than the outcome which would have been achieved had they each cooperated. But this does not mean that they (or their mereological sum) were in any way irrational. I hold that we should say the same thing about the different

time-slices of an agent in an intrapersonal Prisoner's Dilemma. Importantly, I argue not only that we *can* treat the different time-slices of an agent in just the way that we treat the different agents in an interpersonal Prisoner's Dilemma, but that we should be led to do so by an independently motivated account of an agent's options.[22] To this account of options we now turn.

4   Options as Decisions

Determining what you rationally ought to do can be broken down into two stages. The first stage is identifying your options. The second stage is ranking those options. While the second stage has received widespread attention from philosophers of all stripes, the question of what your options are has gone largely unaddressed. But answering this question is crucial to evaluating The No Way Out Argument and hence the import of Diachronic Tragedy.[23]

I am concerned with a *subjective* notion of rationality, which is sensitive to your perspective on the world. What you rationally ought to do in this sense depends on how you believe the world to be, rather than how it actually is, and on what your preferences are, rather than what is objectively valuable. Suppose that you are thirsty and believe the liquid in the glass to be gin, when it is in fact petrol, as in Williams' famous example (Williams (1982)). While there may be a sense in which you ought not reach for the glass, I am concerned with the sense of *ought* in which you ought to reach for the glass. This, in my view, is the sense of *ought* that is central to our conception of rationality.[24]

---

[22]Christensen (1991) likewise emphasizes the parallels between the intrapersonal and interpersonal cases, but his focus is on the Diachronic Tragedy Argument for Reflection in particular, as opposed to Diachronic Tragedy more generally. In considering the Diachronic Tragedy Argument for Reflection, Christensen considers a parallel 'Double Agent' case in which one agent accepts one set of bets and another agent accepts another set of bets, such that all the bets together guarantee the two agents a collective loss. Christensen rightly notes that this in no way impugns either's beliefs. The attitudes of two distinct agents are in no way rationally required to cohere with each other, and Christensen notes that it is tempting to draw the same conclusion about the same agent at different times. I am sympathetic to Christensen's line of thought and especially to his bringing together the interpersonal and intrapersonal cases. My own argument goes further than Christensen's in that it generalizes to all cases of Diachronic Tragedy and provides a principled rebuttal based on a theory of options of a powerful argument for the conclusion that Tragic Attitudes really are in some sense inconsistent.

[23]See also Hedden (2012) for further defense of the conception of options defended in this section.

[24]Some philosophers, such as Thomson (1986), deny that there is this subjective sense of *ought* and hold instead that there is only one sense of *ought* - the objective *ought*. In this sense of *ought*, you ought

Given this subjective notion of rationality, the rational *ought* is supposed to play three theoretical roles. First and foremost, it is to play an evaluative role, in determining whether you and your actions are rationally criticizable or not. Second, it is to play a predictive role; against the background assumption that you are rational, we predict that you will do the thing that you rationally ought to do.[25] Third, the rational *ought* is to be in some sense action-guiding. Now, it is tempting to cash out this action-guiding role by saying that you should always be in a position to know what you rationally ought to do. While I have some sympathy toward this formulation, it is also potentially problematic. For Williamson (2000, ch. 4) has argued that *no* states are luminous, in the sense that whenever they obtain, you are in a position to know that they obtain.[26] It is worth emphasizing that Williamson's argument relies on controversial assumptions about knowledge and hence can be resisted.[27] But I propose to remain neutral on the soundness of Williamson's anti-luminosity argument and refrain from cashing out action-guidingness in terms of always being in a position to know what you rationally ought to do. Instead, we should demand that the rational *ought* be action-guiding in the sense that what you rationally ought to do is sensitive to your uncertainty about the world, so that our theory of the rational *ought* can be of some aid to you in practical deliberation. What you rationally ought to do should depend on what information you have available, rather than simply on how the world in fact is.

These three theoretical roles to be played by the rational *ought* impose, I think, two desiderata on a theory of options. Importantly, I think that each of these theoretical roles independently motivate both desiderata, and so if you object to one of the theoretical

---

not reach for the glass of liquid in Williams' case, since the liquid is in fact petrol. Nevertheless, Thomson concedes that it would be 'sensible' for you to reach for the glass, given that you believe that it contains gin. She just denies that there is a sense of *ought* that lines up with what it would be 'sensible' for you to do. If you side with Thomson in denying that there is a subjective sense of *ought*, you can simply substitute 'it would be sensible for you to ϕ' for each occurrence of 'you rationally ought to ϕ' in what follows.

[25]See e.g., Stalnaker (1984). Note that the predictive and evaluative roles are linked, in that we only predict that you will do whatever it is that you rationally ought to do if we have in place the background assumption that you are rational.

[26]A caveat: As Williamson notes, his argument does not apply to trivial cases of states which always obtain or always fail to obtain, and so these states may in fact be luminous. But this does not threaten the import of his conclusion.

[27]See especially Berker (2008) for a rebuttal of Williamson's argument.

roles but not to another, then you should still accept these desiderata.

Desideratum 1: What your options are supervenes on your present mental states.

Why is this? If what your options are failed to supervene on your present mental states, then what you rationally ought to do would likewise fail to supervene on your present mental states. This would threaten the ability of the rational *ought* to play each of the three theoretical roles outlined above. First, it would make the rational *ought* insufficiently action-guiding. What you rationally ought to do would not be fully sensitive to your perspective on the world. Note also in this regard that on standard frameworks for *ranking* your options, such as expected utility theory, how options are ranked depends only on your present mental states, and it is natural to expect our way of *identifying* your options to likewise do so in a way that depends solely on your present mental states. To the extent that expected utility theory ranks options in a way that allows the rational *ought* to be action-guiding, I think we should expect our theory of what your options are in the first place to place no further obstacles in the way of its ability to be action-guiding. Second, it is implausible to evaluate two agents differently (criticizing the one but praising the other) if they perform the same action after starting out in the same mental states, but this is what would be required if we say that *ought*s fail to supervene on mental states, so that two agents with the same mental states can differ in which sort of action they ought to perform.[28] Third, the use of the rational *ought* in the prediction of action would yield wildly implausible predictions. Suppose you and your doppelgänger, with the same mental states, are each hiking through the forest and come to a raging creek. You are able to ford the creek, but your doppelgänger is not. If options failed to supervene on your mental states (e.g. by consisting of things that you are physically able to do), we could get the result that while you ought to ford the creek, your doppelgänger ought to do something quite different, like give up and head home. But, given that we predict that

---

[28]This is just to express sympathy with a very moderate form of internalism about practical rationality. I support internalism about rationality, both practical and epistemic, largely for the standard reason that it seems inappropriate to criticize an agent for holding some belief or performing some action when the agent was in no position to see that she oughtn't hold that belief or perform that action. Moreover, many motivations for adopting externalism about epistemic rationality (respecting the link between justification and truth and responding to skepticism, for example) do not give rise to parallel motivations for adopting externalism about practical rationality.

rational agents will do what they rationally ought to do, we would then predict that you will ford the creek, while your doppelgänger will do an about-face and head home without so much as getting his feet wet. But this would be bizarre! The two of you are in exactly the same mental state, after all! What could be the source of your radically different behavior? If you immediately performed such very different actions despite being in the very same mental state, it would appear that at least one of you wasn't fully in control of your actions.[29]

Desideratum 2: If ϕ is an option for you, you must be able to ϕ.

Why is this? Desideratum 2 is required to avoid violating the principle that *ought* implies *can*. This principle is important for three reasons. First, you are not subject to any rational criticism for failing to do something which in fact you couldn't have done. Second, *ought implies can* is essential for the predictive role of the rational *ought*, since we would not want to predict that you would do something which in fact you cannot do. Third, the rational *ought* gives you poor guidance if it tells you to do something that you cannot do.[30]

---

[29]Note that denying that you or your doppelgänger ought to ford the creek is not to say that either of you ought not ford the creek; *not-ought* does not entail *ought not*. If ϕ-ing isn't an option for you, then it will be the case neither that you ought to ϕ nor that you ought not ϕ (although on my view, it might be that you ought to *decide* to ϕ).

[30]It is tempting to add a third desideratum to the list, but at it is more problematic the Desiderata 1 and 2, and also not needed to motivate my theory of options, I relegate discussion of it to a footnote. This potential third desideratum says that if ϕ is an option for you, you must *believe* that you are able to ϕ and that ϕ-ing is completely under your control. Why might you think this? Well, to begin with, if you are in fact able to ϕ but believe that you cannot do so, it is implausible to say that nonetheless you rationally ought to ϕ. Suppose you are hiking along and come to a raging creek. You are in fact able to ford the creek (where fording entails that you make it across), and among the things you are actually able to do, fording the creek is best. But you have serious doubts about whether you are able to ford the creek, since it is swollen with the spring snowmelt and flowing fast. Ought you ford the creek? Intuitively, I think not.

Another way to motivate this third desideratum is this: Standard frameworks for ranking options (e.g., expected utility theory) have the feature that their evaluation of fording, say, cannot take into account the potential bad consequences of trying to ford but failing to do so. This is because the expected utility of fording is the sum of the utilities of possible outcomes, weighted by the probability that that outcome would obtain, given that you ford the creek. Therefore, the utility of an outcome in which you try but fail to ford the creek makes no difference to the expected utility of fording, since the probability of an outcome in which you try to ford but fail, given that you ford the creek, is zero! But your beliefs about what would happen if you were to try to do something but fail in this attempt are often very important for your own deliberations and for our evaluation of you. This potential third desideratum is meant to mitigate the impact of this feature of expected utility theory (that its evaluation of an action does not take into account consequences of trying but failing to perform that action) by restricting your options to things that you believe you are able to do.

What sort of theory options could satisfy these two desiderata? While it might at first be tempting to think of your options as all of the actions you are able to perform,[31] this proposal violates Desideratum 1, since you might be able to perform some action without its being the case that any other agent in the same mental state would likewise be able to perform that action (as in the case of your doppelgänger). Which actions you are able to perform does not supervene on your present mental states.

A second proposal might be that your options consist of all and only the actions you *believe* you are able to perform. But while this proposal satisfies Desideratum 1, it violates Desideratum 2, since you might falsely believe of some action that you are able to perform it.

A third proposal might be to take your options to consist of all and only the actions that you *know* you are able to perform. This proposal clearly satisfies Desideratum 2, since knowledge is factive. Does it satisfy Desideratum 1? That depends on whether knowledge is a mental state. Williamson (2000, ch. 2) argues that it is. His argument proceeds primarily by rejecting reasons for thinking that knowledge is not a mental state. First, one might think that knowledge isn't a mental state, since whether you know a proposition depends on factors external to your physical state (e.g., on whether the proposition is true). But this is true of mental states more generally, since the content of your attitudes is determined by factors external to your physical state. Second, one might think that knowledge isn't a mental state because you are not always able to tell whether you know, as opposed to merely believe, a proposition. But if Williamson's anti-luminosity argument (ch. 4) is sound, no states - mental or non-mental - are such that whenever they obtain, you are in a position to know that they obtain.

If Williamson is right, then this third proposal - that your options are all and only the actions you know you are able to perform - does indeed satisfy both of our desiderata. And for present purposes, I want to remain neutral on whether knowledge is in fact a

---

There is, however, a worry about whether there are *any* actions (even mental acts of decision-making) which you are certain are completely under your control. So this potential third desideratum, though well-motivated, might be too strict to be satisfiable.

[31]It seems that Jeffrey (1965) and Lewis (1981), among others, held versions of this theory of options.

mental state.[32]

But I think that there are in fact good grounds for strengthening Desideratum 1 so that options must supervene not only on your mental states, but on your *non-factive* mental states, to yield:

> **Desideratum 1***: What your options are supervenes on your present non-factive mental states.

Why adopt this strengthened version of Desideratum 1? What is special about non-factive mental states? There may be theoretical reasons for privileging non-factive states in theorizing about rationality. For instance, Wedgwood (2002) argues that where a factive attitude is constituted by a non-factive attitude, as in the case of knowledge and belief, rational requirements should make reference to the latter, since the latter will figure in more proximal explanations of belief and behavior. Wedgwood's argument for the centrality of non-factive mental states is worthy of consideration, but I think that there is a simpler, less theoretical motivation for strengthening Desideratum 1 to Desideratum 1*.

Consider two cases, one in which you know you are able to ϕ, and one in which you are in a Gettier situation and hence only justifiably believe the true proposition that you are able to ϕ. Intuitively, I think, the facts about what you rationally ought to do are the same in both cases; either you ought to ϕ in both cases or you ought not ϕ in both cases.[33] Consider the following cases (which are somewhat analogous to the cases of you and your doppelgänger considered above):

> Case 1: You are hiking through the forest and come to a raging creek. You must either ford it or turn back. Your hiking partner, who knows how deep the creek is, knows that you are able to ford it and tells you that you are able to do so. On the basis of this testimony, you gain knowledge that you are able to ford the creek.

---

[32]See Fricker (2009) for the opposing view that knowledge is not a mental state.

[33]Note that we cannot get around this issue by saying that your options are all and only the actions you truly believe you are able to perform. For this proposal fails to satisfy Desideratum 1. Which actions you truly believe you are able to perform does not supervene on your mental states, regardless of whether we count factive attitudes like knowledge as mental states.

Case 2: Just as in Case 1, except your hiking partner in fact doesn't know how deep the creek is, but instead merely confidently asserts it. On the basis of this testimony, you gain the justified true belief that you are able to ford the creek. But because your hiking partner did not know this that you are able to ford it, your justified true belief that you are able to ford the creek does not constitute knowledge.[34]

In Case 1, you know (i) that you are able to ford the creek, and (ii) that you are able to turn back. In Case 2, you know (ii) but not (i), since your belief that you are able to ford the creek does not constitute knowledge, despite being true and justified. So if your options are all and only the actions you know you are able to perform, then fording the creek is among your options in Case 1 but not in Case 2. Assuming that you much prefer getting across to turning back, this means that we will get the result that you rationally ought to ford the creek in Case 1 but not in Case 2.

This is, to my mind, a highly problematic result. Return to the evaluative and predictive roles of the rational *ought*. By hypothesis, fording the creek is your best option in Case 1. And the only difference between the cases if that your justified true belief that you can ford the creek constitutes knowledge in Case 1 but not in Case 2; the cases are otherwise identical in all physical and mental respects. I think, therefore, that if in Case 2 you turned around and headed back instead of fording the creek, we would be inclined to judge you harshly. We would deem you highly irrational. Moreover, we would naturally predict that insofar as you will go ahead and ford the creek in Case 1, you will also do so in Case 2.[35]

One might object that the above argument assumes that if in Case 2 fording the creek isn't an option, then turning back is the best option. But even if fording the creek isn't

---

[34]This assumes that testimony only yields knowledge of the testified-to proposition if the testifier knows that proposition. See Fricker (2006, 602) for a defense. If you object to this view of testimony, substitute another way of giving you a Gettier-ized belief that you are able to ford the creek.

[35]Of course, as Williamson (2000, ch. 2) notes, the fact that your justified true belief that you can ford the creek is Gettier-ized in Case 2 means that you are more likely to encounter defeating evidence down the road. Perhaps your friend will have a change of heart and confess his ignorance, whereas your friend in Case 1 is unlikely to inexplicably lie and claim ignorance. But the point is that we would predict that you would at least start off in the same way in Case 1 and Case 2, even if the cases differ in terms of what evidence you might encounter later on.

an option, since you don't know you are able to do so, maybe *trying* to ford the creek is still an option. After all, even though you don't know that you are able to successfully ford the creek, you presumably do know that you can try. And given that you want to get across and believe you are able to do so, presumably *trying* to do so looks better than just giving up and heading back. So while we would indeed evaluate you harshly if you just turned around, this doesn't require that fording the creek count as an option for you, but only that trying to ford counts as an option. And similarly, we still get the result that we would predict that you would try to ford.

However, once we move to evaluating tryings, it seems to me that we should simply abandon the proposal that your options are the actions you know you are able to perform (which may include tryings), and instead say that your options are simply the tryings themselves. This is the approach I advocate.

Of course, we might ordinarily think of tryings as physical actions, so that trying to ford the creek actually requires wading in, for instance. But on this conception of tryings, which things you can try to do will fail to supervene on your mental states, just as which ordinary actions (non-tryings) you can perform will fail to supervene on your mental states. For instance, whether you are able to wade into the water will depend not just on your mental states but also on whether your legs are working, whether your shoelaces are tied together, and the like. But if tryings are understood as mental actions which start the ball rolling, so to speak, then I am happy to think of your options as tryings. To enforce this reading, I will call these mental tryings 'decisions.' This approach yields the following theory of options:

> **Options-as-Decisions**: Your options at t consist of all and only the actions
> of the form *You decide at t to ϕ* that you are able at t to perform.

So, instead of the rational *ought* applying to ordinary non-mental actions like *fording the creek*, it will only apply to mental volitional acts of making decisions.

But does **Options-as-Decisions** satisfy Desiderata 1* and 2? It clearly satisfies Desideratum 2; this is built into the wording of the proposal. So the question is whether

it satisfies Desideratum 1*. It is the case that which decisions you are able to make at a time supervenes on your mental states at that time?

The answer to that question depends on what it takes to be able to make a given decision. I do not want to be fully committal about what it takes to be able to make a decision, but I think that on any attractive theory of decisions, which decisions you are able to make will supervene on your mental states in a way that allows **Options-as-Decisions** to satisfy Desideratum 1*.

Consider for instance the account of Bratman (1987), to which I am sympathetic. He holds that you are able to decide (or intend) to ϕ just in case you do not believe that, were you to decide to ϕ, you would not ϕ. You can make a decision so long as you do not believe that you wouldn't carry it out. Bratman's account is supported, *inter alia*, by the Toxin Puzzle (Kavka (1983)).[36] In the Toxin Puzzle, you are presented with a drink containing a mild toxin which will cause you moderate temporary discomfort but will not result in any long-term harm. You are offered a large sum of money if at midnight tonight you decide to drink the toxin tomorrow afternoon. You do not actually have to drink the toxin to gain the reward; you only have to make the decision at midnight to do so. It seems that despite the benefits of deciding to drink the toxin, you are unable to do so, for you realize that were you to decide at midnight to do so, you would promptly reconsider and overturn your earlier decision. You cannot decide the drink the toxin because you believe that, were you to decide to do so, you would not carry out this decision. If this is the only restriction on which decisions you are able to make, we get Bratman's account. And if Bratman is right, then **Options-as-Decisions** satisfies Desideratum 1*, since it is your beliefs which determine which decisions you can make.

Other accounts of abilities to decide will also allow **Options-as-Decisions** to satisfy Desideratum 1*. Whether you are able to make some decision might depend not only on your beliefs, but also on your desires, so that e.g., you are unable to decide to do something to which you have an extremely strong aversion. But if this aversion is a desire

---

[36]More carefully, Bratman and Kavka are both concerned with intentions, but insofar as there is a difference between decisions and intentions, it seems that the considerations they raise in this regard apply equally to both.

or a fear or otherwise part of your mental state, then this is still compatible with the claim that **Options-as-Decisions** satisfies Desideratum 1*. Other sorts of psychological pathologies might also impact which decisions you can make, but again, if these psychological pathologies are part of your mental state, my account of options is still in good shape. (Of course, psychological pathologies might not be propositional attitudes, but I am conceiving of your mental state as including both attitudinal and non-attitudinal states.)[37] So while it is a difficult question what it takes in order to be able to make a given decision, and while I do not advocate any particular answer to this question, I think that it is rather plausible that which decisions you are able to make supervenes on your mental states, and hence that **Options-as-Decisions** satisfies both of our desiderata.[38]

Summing up, there are good reasons to think that your options should supervene on your (non-factive) mental states and be things that you are able to do. A proposal on which your options are all and only the actions you know you can perform satisfies these desiderata but is otherwise problematic. Instead, we should adopt **Options-as-**

---

[37]Thanks to an anonymous referee for suggesting this possibility. In the referee's case, two otherwise identical agents are such that one can make decision D while the other is unable to make decision D due to some psychological pathology. This is compatible with which decisions they are able to making supervening on their mental states, even though which decisions they can make might not supervene on a more narrow class of mental states, such as their beliefs and desires. Or, for a non-pathological case, it might be that decision D would involve cruelty, and one of the agents is unable to be cruel. Now, this might be akin to the Toxin Puzzle, where the one cannot make the decision because carrying out the decision is cruel, and so she believes she would not carry out the decision no matter what. In this case, it is her beliefs, and so her mental state more broadly, which determines whether she can make that decision. Or it might be that she has no beliefs about whether she would carry out the decision but is simply unable to make the decision due to her aversion to cruelty. But again, as long as this aversion is part of her mental state (whether because it is a desire or whether it is some other kind of attitude or non-attitudinal state), **Options-as-Decisions** can still satisfy Desideratum 1*.

[38]What if your abilities to make decisions are restricted not just by your own mental states, but also by external forces? Frankfurt (1969) considers the possibility of a demon who monitors your brain activity and will strike you down if you are about to decide to ϕ. Plausibly, you are unable to decide to ϕ. The possibility of such a demon threatens the claim that which decisions you are able to make supervenes on your mental states. It also threatens the claim that you are always in a position to know which decisions you are able to make, since you may not be in a position to know whether such a demon is watching you.

I am inclined to say that in such a case, where a Frankfurtian demon is monitoring you in this way, there may not be *anything* that you rationally ought to do. Suppose that, among the decisions you believe you are able to make, the decision to ϕ looks best, but a demon is going to strike you down if you are about to decide to ϕ. What ought you to do in this case? Certainly, it is not that you ought to make some decision other than the decision to ϕ, since all such decisions look inferior by your lights. And it is not the case that you ought to decide to ϕ, since *ought* implies *can*. The right thing to say is that there is nothing that you ought to *do*; rather, you ought to be in a state of being about to decide to ϕ, where this will lead to your being struck down before you are actually able to *do* anything at all. The rational *ought* thus only applies to agents who are not being disrupted by Frankfurtian demons in this way, and so once we restrict our attention to agents to whom the rational *ought* applies, which options an agent has will both supervene on her beliefs and desires and be knowable by her.

Decisions, on which the requirements of rational action apply only to particular decisions made by particular time-slices of the agent.[39] And as we shall now see, this conception of an agent's options entails the falsity of P2 of The No Way Out Argument.

5 Decisions and Diachronic Tragedy

According to **Options-as-Decisions**, your options (the things to which the rational *ought* applies) at a time consist of all and only the decisions you are able to make at that time. (These decisions can have contents of very different types; they can be decisions to perform particular actions, decisions to perform sequences of actions,[40] decisions to defer deliberation until a later time, etc.) This means that evaluating whether you acted rationally in performing some sequence of actions requires us to think carefully about what was going on in your head and about the individual decisions you made on the way to performing that sequence of actions. Evaluating whether you were rational over a period of time requires determining whether you were rational in making the decisions you made at each particular time.

Recall P2 of The No Way Out Argument:

---

[39]An anonymous referee raises the worry that by 'retreating into the mental,' so to speak, **Options-and-Decisions** threatens the ability of the rational *ought* to play the predictive role outlined above. It would allow us to predict which decision an agent will make (again, against the background assumption that she is rational), but not what she will do from there. But I think that this is acceptable. Our theory of the rational *ought* yields one particular prediction - namely that the agent will take the option with highest expected utility - and then our general knowledge of the case allows us to make further predictions from there. For example, if we, the theorists, know that there are no obstacles to the agent's carrying out the decision she makes, then we will predict that she will in fact go on to carry out the decision. This is because it is part of the functional role of decisions that they tend to cause their being carried out. And if we know what certain other consequences of her carrying out that decision will be, then we will predict that those consequences will also occur. Take once more the case of the raging creek. Expected utility theory allows us to predict that you will decide to ford the creek (or perhaps decide to try to do so). If we know that your shoelaces are not tied together, that your legs are strong enough and the creek not too deep, that you do not tend to get scared at the last minute, and the like, then we will predict that you will in fact ford the creek, even though this prediction isn't given by expected utility theory alone. This phenomenon holds regardless of what theory of options we accept. In general, expected utility theory yields one specific prediction about what will happen, but we can supplement this prediction with other background knowledge that we possess to generate further predictions.

[40]Indeed, there may not be a principled distinction between particular actions and sequences thereof. The vast majority of the non-mental actions we perform, such as buying groceries, can be thought of either as actions in their own right or as sequences of many smaller actions, such as stepping out the door, walking to the store, putting fruit in the basket, etc. A further attractive feature of **Options-as-Decisions** is that it does not require arbitrary stipulations about the fineness of grain of the actions which should be evaluated in deliberation. All possible decisions count as options, no matter the fineness of grain of their contents.

P2: If you have Tragic Attitudes, then in some cases no matter what you do you will have done something that, given those attitudes, you rationally ought not have done.

How does P2 fare in light of **Options-as-Decisions**? Supposing you perform a Tragic Sequence, have you done anything that you rationally ought not have done? The first thing to note is that given **Options-as-Decisions**, we cannot immediately infer that you have done something you rationally ought not have done by performing a Tragic Sequence, since sequences of actions are not among your options (although, importantly, *decisions* to perform certain sequences of actions will be among your options; it's just that the sequences themselves won't count as options).[41] Instead, we must consider how your performing a Tragic Sequence came about, and in particular whether it was the result of any *decision* that you rationally ought not have made.

And crucially, we can specify the details of cases of Diachronic Tragedy in such a way that you perform a Tragic Sequence without ever making a decision that you rationally ought not have made, given your beliefs. Here is one such way: You did fully believe that you would carry out whichever decision you made (i.e. you believed yourself to be able to self-bind), but you were in fact wrong about this. Suppose that in your youth, you made the decision to perform the sequence <Donate Early, Not Donate Late>. Given your belief that you would do whatever you decided to do, this was the decision you rationally ought to have made (since you preferred this sequence of actions over any other). But despite having made this decision in your youth and carried out the first part of it (Donating Now), you reopened deliberation later on in life and revised the decision you made in your youth, deciding instead to Donate Late. This new decision was also one that (having reopened deliberation) you rationally ought to have made, since your older self preferred Donating Later to Not Donating Later. Having carried out this

---

[41]Consider a sequence of decisions consisting of decision D_1 at t_1 and decision D_2 at t_2. Thinking back to Desiderata 1 and 2, it may be that you are able to make each decision at the relevant time. And it may be that whether you are able to make decision D_1 at t_1 supervenes on your mental state at t_1 and that whether you are able to make decision D_2 at t_2 supervenes on your mental state at t_2, but crucially whether you are able to perform the sequence $< D_1$ at t_1, D_2 at $t_2 >$ does *not* supervene on your mental state at t_1. This is why this sequence of decisions should not count as an option for you at t_1. Whether you can perform this sequence of decisions depends not just on your mental states at t_1 but also on facts about how things will be in the future.

new decision and Donated Later, you wound up performing the Tragic Sequence <Donate Early, Donate Late>. But at no point in the process did you take an option (that is, make a decision) that you rationally ought not have taken, given your beliefs and preferences. Your performing the Tragic Sequence was the result not of having made any decision that you rationally ought not have made, but of having falsely believed that you would carry out whichever decision you made.

Here is a second way: Your performing a Tragic Sequence was the result of failing to believe that you would carry whichever decision you made (i.e. failing to believe yourself able to self-bind). Suppose that in your youth, you believed that your present decision would make no difference to which action you perform at age 60; it would only determine whether you would Donate Early or not. In this case, three decisions were tied for best: (i) the decision to perform the sequence <Donate Early, Donate Late>, (ii) the decision to perform the sequence <Donate Early, Not Donate Late>, and (iii) the decision to Donate Early (deferring until age 60 the question of whether to then Donate Late). Suppose you make the third decision and carry it out. This decision was perfectly rational in light of your beliefs and preferences. Then, at age 60 you must deliberate anew and decide to Donate Late. This new decision was also perfectly rational, since at that point you preferred to Donate Late. You wind up performing the Tragic Sequence <Donate Early, Donate Late>, but at no point did you make a decision that you rationally ought not have made. Your decision in your youth to Donate Early was perfectly rational (given your belief that you could not control your 60 year old self), as was your decision at age 60 to Donate Late. Your performing the Tragic Sequence was the result not of having made a decision that you rationally ought not have made, but of having failed to believe that you would carry out whichever decision you made.

In sum, given **Options-as-Decisions**, P2 is false. To begin with, we cannot immediately infer from your performing a Tragic Sequence that you did something you rationally ought not have done, since sequences of actions are not options; they are not the sorts of things to which the rational *ought* applies. Whether your performing a Tragic Sequence involved your doing anything you rationally ought not have done depends on whether

the individual decisions leading to that Tragic Sequence were themselves irrational. But in a wide range of cases your performing a Tragic Sequence is instead be the result of a sequence of perfectly rational decisions. So P2 is false and the best argument that Tragic Attitudes are irrational, The No Way Out Argument, fails.

Let's return to the Prisoner's Dilemma. According to **Options-as-Decisions**, the rational *ought* applies to particular decisions that can be made by particular time-slices of an agent. Mereological sums of time-slices and sequences of actions are not the proper subjects and objects, respectively, of the rational *ought*. As a result, we cannot infer from Diachronic Tragedy that the preferences of the particular time-slices involved are irrational. Cases of Diachronic Tragedy are essentially *intra*personal Prisoner's Dilemmas, and I am in effect arguing that **Options-as-Decisions** should motivate us to say about cases of Diachronic Tragedy the same thing that is standardly said (e.g., by Parfit (1984)) about the *inter*personal Prisoner's Dilemma. In the latter, each prisoner prefers to defect, but by each defecting, they wind up with an outcome which by each of their lights is worse than the outcome which would have been achieved had they each cooperated. But this does not mean that they performed an irrational 'group action' or that their mereological sum was irrational, since group actions and mereological sums of people are not proper relata of the rational *ought*. And it does not mean that either individual's preference for defecting over cooperating was irrational. While treating the time-slices involved in cases of Diachronic Tragedy in the same way that we already treat the people involved in the Prisoner's Dilemma may seem radical, I am arguing that we are led to do so by an independently motivated account of options.

Of course, there is still a crucial difference between the intrapersonal and interpersonal Prisoner's Dilemmas which might be thought to undermine my argument. Whereas the two prisoners in the standard Prisoner's Dilemma do not care about each other's well-being, you probably *do* care a great deal about the well-being of your future selves. You may even be rationally *required* to care about the well-being of your future selves. Does this sort of self-concern undermine my claim that intrapersonal and interpersonal Prisoner's Dilemmas should be treated alike? No, it does not. To say that you are

rationally required to care about your future well-being is just to say that you ought now have certain preferences; namely that you ought to prefer, *ceteris paribus*, that you be better rather than worse off in the future. It is *not* to say that your present and future selves are rationally required to coordinate their actions in the manner necessary to avoid suffering Diachronic Tragedy. This would be a separate requirement that goes well beyond the requirement that you prefer that your future selves be well-off. And importantly, none of the cases of Diachronic Tragedy assumed that you lacked appropriate concern for your future well-being. For instance, in the case of **Satan's Apple**, we can assume that you care deeply about the enjoyment of your future selves. That is why you want to amass as much apple as possible while avoiding eternal damnation. You suffer Diachronic Tragedy not because of a lack of self-concern, but rather because, acting on this self-concern each time a slice of apple is offered leads to the horrible outcome of your eating infinitely many slices of apple and being sent to Hell. Therefore, in claiming that cases of Diachronic Tragedy should be treated the same as interpersonal Prisoner's Dilemmas, I am not denying that you ought to be concerned for your future well-being. Rather, I am denying that this rational requirement of self-concern amounts to the claim that you are rationally required to be invulnerable to Diachronic Tragedy. Instead, this rational requirement of self-concern should be understood as a requirement that you now prefer that you be well-off in the future, where having this preference is compatible with your rationally suffering Diachronic Tragedy.

6   Is Everything then Permitted?

If **Options-as-Decisions** is a proper characterization of your options, then we are left with no compelling argument that Tragic Attitudes are irrational. I draw the stronger conclusion that the fact that certain attitudes yield Diachronic Tragedy in no way entails that those attitudes are irrational.

Are we therefore committed to thinking that all of the attitudes discussed in Section 2 are perfectly rational? Not at all. In some cases, the attitudes in question are in fact irrational, albeit for reasons entirely independent of Diachronic Tragedy.

One might, for instance, argue for the irrationality of intransitive preferences along the following lines, loosely inspired by Kolodny.[42] A preference for one thing over another is supported by reasons only if there is more reason to desire the one than to desire the other. But it is a metaphysical truth that 'more,' and hence 'more reason than,' is transitive, and so there cannot be more reason to desire A than to desire B, more reason to desire B than to desire C, and more reason to desire C than to desire A. Therefore, if you have intransitive preferences (e.g. by preferring A to B, B to C, and C to A), the structure of your preferences itself entails that not all of your preferences can be supported by reasons. But if one can read off from the mere structure of your preferences that they cannot all be supported by reasons, then those preferences are irrational. Hence it is irrational to have intransitive preferences.[43]

While this sketch of an argument will of course be controversial, the important thing to note is that there are promising ways of arguing that certain attitudes traditionally targetted by appeal to Diachronic Tragedy are in fact irrational, albeit for reasons entirely independent of Diachronic Tragedy.

But with other cases of Tragic Attitudes, it is doubtful whether there will be any such independent argument that they are irrational, and we should instead conclude that the attitudes are rational despite giving rise to Diachronic Tragedy. This is likely the case, for instance, with the preferences at issue in **Satan's Apple**. All of your preferences are perfectly defensible and coherent, but infinities are strange beasts, and even perfectly rational preferences can land you in trouble if you face infinitely many decision points. As for the other cases, like preference shifts, time-bias, and imprecise preferences, whether we ought to conclude that they are irrational will likewise depend on whether there is some *independent* argument for their irrationality, perhaps along the lines of the Kolodny-inspired argument against intransitive preferences.

This fact could potentially explain the curious sociological fact mentioned in Section 3, namely that philosophers have been willing to conclude from *some* instances of Diachronic

---

[42]See especially Kolodny (2007) and Kolodny (2008).

[43]A slightly different argument would appeal to a link between rational preferences and betterness, along with the metaphysical or axiological claim that betterness is transitive.

Tragedy that the relevant attitudes are irrational but have been reluctant to draw the same conclusion in other instances of Diachronic Tragedy. It may be that in the instances where philosophers have been willing to use Diachronic Tragedy to argue for the irrationality of the attitudes involved, there is an independent Kolodny-style argument that those attitudes are irrational, whereas in the instances where philosophers have been reluctant to draw the same conclusion, this is because there is no such independent argument for the irrationality of the relevant attitudes. Perhaps intuitions about rationality or irrationality in cases of Diachronic Tragedy are really tracking whether there is some such independent argument that could support the charge that the attitudes involved are irrational.

But whether or not this psychological speculation about the source of our intuitions is correct, it remains the case that we *ought* to evaluate the rationality of Tragic Attitudes on independent grounds; just pointing out that they yield Diachronic Tragedy is not enough.

7   Conclusion

There are myriad cases where having certain attitudes predictably leads you to perform sequences of actions that are to your own acknowledged disadvantage. They yield Diachronic Tragedy. But the attitudes themselves have little else in common. Some philosophers have placed great weight on this single shared feature and concluded that certain Tragic Attitudes are irrational. I have argued that on a proper conception of what your options are, the best *argument* for this conclusion fails, and that we have no reason to think that Tragic Attitudes are *ipso facto* irrational.

This single shared feature of the attitudes discussed in Section 2 should not blind us to their many differences. In some cases (e.g. intransitive preferences), the attitudes really are irrational, but this irrationality is identifiable independently of any considerations about sequences of actions or predictable exploitability. But in many other cases, the attitudes are perfectly rational. They are defensible and supported by reasons at all times despite their potential to lead to misfortune.

Diachronic Tragedy is a symptom. But like many symptoms, it can result from a variety of different diseases, or even in the absence of any disease at all. In some cases,

Diachronic Tragedy is the symptom of irrational attitudes, but in others, Diachronic Tragedy arises in the absence of any irrationality at all. Therefore, to give a proper diagnosis of a case of Diachronic Tragedy, we must look beyond this superficial symptom and direct our attention to the underlying attitudes themselves. We must determine on independent grounds whether the attitudes are defensible and supported by reasons, rather than merely noting that they yield Diachronic Tragedy.[44]

References

Anscombe, Elizabeth. 1957. *Intention*. Oxford: Basil Blackwell.

Arntzenius, Frank, Adam Elga, and John Hawthorne. 2004. Bayesianism, Infinite Decisions, and Binding. *Mind* 113:251–283.

Berker, Selim. 2008. Luminosity Regained. *Philosophers' Imprint* 8:1–22.

Bratman, Michael. 1987. *Intentions, Plans, and Practical Reason*. CSLI.

———. 2007. Reflection, Planning, and Temporally Extended Agency. In *Structures of Agency*. Oxford University Press.

Briggs, Rachael. 2009. Distorted Reflection. *Philosophical Review* 118:59–85.

Christensen, David. 1991. Clever Bookies and Coherent Beliefs. *Philosophical Review* 100:229–247.

———. 1996. Dutch-Book Arguments Depragmatized: Epistemic Consistency for Partial Believers. *Journal of Philosophy* 93:450–479.

Davidson, Donald, J. C. C. McKinsey, and Patrick Suppes. 1955. Outlines of a Formal Theory of Value, I. *Philosophy of Science* 22:140–160.

Dougherty, Tom. 2011. On Whether to Prefer Pain to Pass. *Ethics* 121:521–537.

———. ms. A Deluxe Money Pump Argument .

Elga, Adam. 2010. Subjective Probabilities Should be Sharp. *Philosopher's Imprint* 10.

Frankfurt, Harry. 1969. Alternate Possibilities and Moral Responsibility. *Journal of Philosophy* 66:829–839.

Fricker, Elizabeth. 2006. Second-Hand Knowledge. *Philosophy and Phenomenological Research* 73:592–8211.

———. 2009. Is Knowing a Mental State? The Case Against. In *Williamson on Knowledge*. Oxford University Press.

Hedden, Brian. 2012. Options and the Subjective Ought. *Philosophical Studies* 158:343–360.

———. forthcoming. Incoherence Without Exploitability. *Noûs* .

Hinchman, Edward. 2003. Trust and Diachronic Agency. *Noûs* 37:25–51.

Jackson, Frank, and Robert Pargetter. 1986. Oughts, Options, and Actualism. *Philosophical Review* 95:233–255.

Jeffrey, Richard. 1965. *The Logic of Decision*. University of Chicago Press.

Joyce, James. 1998. A Nonpragmatic Vindication of Probabilism. *Philosophy of Science* 65:575–603.

Kavka, Gregory. 1983. The Toxin Puzzle. *Analysis* 43:33–36.

Kennedy, Ralph, and Charles Chihara. 1979. The Dutch Book Argument: its Logical Flaws, its Subjective Sources. *Philosophical Studies* 36:19–33.

Kolodny, Niko. 2007. IX-How Does Coherence Matter? *Proceedings of the Aristotelian Society* 107:229–263.

———. 2008. Why Be Disposed to Be Coherent? *Ethics* 118:437–463.

Korsgaard, Christine. 2008. *The Constitution of Agency*. Oxford University Press.

Lewis, David. 1981. Causal Decision Theory. *Australasian Journal of Philosophy* 59:5–30.

———. 1999. Why Conditionalize? In *Papers in Metaphysics and Epistemology*. Cambridge University Press.

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

Raiffa, Howard. 1968. *Decision Analysis*. Wesley-Addison.

Rosenkranz, Roger. 1981. *Foundations and Applications of Inductive Probability*. Ridgeview Press.

Skyrms, Brian. 1987. Dynamic Coherence and Probability Kinematics. *Philosophy of Science* 54:1–20.

Stalnaker, Robert. 1984. *Inquiry.* Cambridge University Press.

Taylor, Shelley, and Jonathon Brown. 1988. Illusion and Well-Being: A Social-Psychological Perspective on Mental Health. *Psychological Bulletin* 103:193–210.

Thomson, Judith Jarvis. 1986. *Rights, Restitution, and Risk.* Harvard University Press.

van Fraassen, Bas C. 1984. Belief and the Will. *Journal of Philosophy* 81:235–256.

Wedgwood, Ralph. 2002. Internalism Explained. *Philosophy and Phenomenological Research* 65:349–369.

Williams, Bernard. 1982. *Moral Luck.* Cambridge University Press.

Williamson, Timothy. 2000. *Knowledge and its Limits.* Oxford University Press.