

# The Tragedy of the Risk Averse\*

H. Orri Stefánsson

[orri.stefansson@philosophy.su.se](mailto:orri.stefansson@philosophy.su.se)

December 19, 2020

## Abstract

Those who are risk averse with respect to money, and thus turn down some gambles with positive monetary expectations, are nevertheless often willing to accept bundles involving multiple such gambles. Therefore, it might seem that such people should become more willing to accept a risky but favourable gamble if they put it in context with the collection of gambles that they predict they will be faced with in the future. However, it turns out that when a risk averse person adopts the long-term perspective, she faces a decision-problem that can be analysed as a noncooperative game between different “time-slices” of herself, where it is in the interest of each time-slice (given its prediction about other slices) to turn down the gamble with which it is faced. Hence, even if a risk averse but rational person manages to take the long-term perspective, she will, in the absence of what Hardin called “mutual coercion”, end up in a situation analogous to the “tragedy of the commons”.

**KEYWORDS:** *Risk aversion; Rationality; Decision framing; Prisoner’s Dilemma.*

---

\*This paper was presented at the PPE seminar at the Institute for Futures Studies in Stockholm, where I received very useful comments and suggestions. Special thanks to Richard Bradley, Krister Bykvist, Alexander Heape, and Johanna Thoma for helpful discussions about the content of this paper, and to two referees and an editor for very useful suggestions. Financial support from Riksbankens Jubileumsfond is gratefully acknowledged.

# 1 Introduction

Imagine being offered a 50-50 gamble between winning \$2,000 and losing \$1,000. Would you accept the offer? If you are risk averse, as many people are (even when relatively little is at stake), then the answer might be “no”. The risk of losing \$1,000 is simply not worth an equal chance of gaining \$2,000. What if instead you were offered a bundle consisting of  $n$  such gambles? If you believe that the chance that each gamble results in a win is independent of the outcomes of the other gambles, then there will presumably be some (finite)  $n$  such that you would accept the bundle, even though you turned down the single gamble. After all, since each gamble has a positive monetary expectation, probabilistic independence between the outcomes of the gambles means that as  $n$  becomes larger, the probability that the bundle results in a monetary loss becomes arbitrarily small while the expected monetary return becomes arbitrarily large.

An attitude like that described above, that is, a preference (at any level of wealth) against a particular gamble combined with a preference in favour of a bundle consisting of some finite number of such gambles, is known to be inconsistent with expected utility theory, as the theory is typically formulated<sup>1</sup> (Samuelson 1963).<sup>2</sup> So, those who believe that the theory correctly captures the requirements of instrumental rationality should think that the combined preference is irrational. In contrast, those who favour alternatives to expected utility theory that can account for attitudes like that described above—for instance, those who favour “risk-weighted” expected utility theory (Buchak 2013)—might take the combined preference to support their favoured theory.<sup>3</sup>

Although I believe that the combined preference is rationally permissible, I shall not in this paper argue that it is.<sup>4</sup> Instead, I address a puzzle that

---

<sup>1</sup>That is, assuming that utilities are defined on terminal outcomes rather than changes.

<sup>2</sup>More precisely, the attitude in question is inconsistent with a combination of Transitivity and Event-Wise Dominance (Stefánsson 2018), given the assumption mentioned in the last footnote.

<sup>3</sup>Note however that Thoma (2019) has argued that in fact, risk-weighted expected utility theory cannot account for the attitude under discussion. Since my aim in this paper is not to discuss the advantages of different decision theories I will not discuss further which theories can, and which cannot, account for the attitude in question.

<sup>4</sup>For recent discussions of the rational permissibility of the attitude under discussion,

arises for those of us who believe that the attitude in question is rationally permissible.

The puzzle derives its practical force from the fact that although most of us are not fortunate enough to be repeatedly offered simple gambles like those discussed above, we are repeatedly faced with more complex independent “gambles”—that is, risky prospects—with positive *objective*<sup>5</sup> expectations. That is, throughout the course of our lives, we face many gambles that have some chance of resulting in a desirable outcome and some risk of resulting in an undesirable outcome, where for each gamble, the probability weighted desirable quantity (e.g. money won, or lives saved) outweighs the probability weighted undesirable quantity (e.g. money or lives lost). So, if one has a combination of preference like that discussed above, then although one might be tempted to reject a gamble when considered in isolation, one might be willing to accept the gamble when viewed in the context of the gambles that one predicts one will be faced with in the future. In other words, whether one accepts a gamble might then depend on whether one takes the long-term or the short-term perspective.<sup>6</sup>

An important ingredient in the said puzzle, is the assumption that it is typically unreasonable, perhaps even irrational, to take the short-term perspective. It is more or less uncontroversial that when framing a decision-problem, one should include everything that might affect the values of, and thus the choice between, the available options. Therefore, if the gambles that one believes one will be faced with in the future affect the values of one’s current options, then one should take this into account, when deciding what choice to make in one’s current situation. But if one accepts this line of reasoning, then it might seem that it is not rationally permissible to turn down the gambles that one *would* accept if one were to take the long-term perspective.<sup>7</sup>

---

see e.g. Buchak (2013), Stefánsson (2018) and Thoma (2019).

<sup>5</sup>By a gamble with a positive *objective* expectation, I mean for instance a bet with a positive monetary expectation, or a public policy whose expected number of lives saved is positive, etc. Such gambles can however of course have a negative expected *utility* (compared to their alternatives, e.g. the status quo).

<sup>6</sup>For evidence that the long vs. short-term perspective affects actual decision-makers’ willingness to accept gambles, see e.g. Thaler et.al. (1997).

<sup>7</sup>It might be worth emphasizing that whenever I speak of “rationality” in this paper, I have in mind the (thin) conception of *instrumental* rationality, according to which no

In sum, the puzzle consists in the seeming tension between the following claims. First, one can rationally permissibly turn down a gamble when viewed in isolation, even though one would accept it if viewed as part of the collection of gambles that one predicts one will be faced with in the future. Second, when evaluating a gamble, one (rationally) should include everything that is relevant to the gamble's choice-worthiness, including the gambles that one predicts one will be faced with in the future. From the second claim it seems to follow that if one would accept a particular gamble when one has taken into account one's beliefs about the future, then it is not rationally permissible to turn down the gamble. But how is that consistent with the first claim?

The aim of this paper is respond to the above puzzle. I shall argue that even if one is rationally required to take into account one's beliefs about the future when evaluating a gamble, a rational person may well turn down a gamble that they would accept if it were offered as part of a bundle consisting of the gambles that the person predicts she will be offered in the future. In fact, I shall argue that for a minimally rational person with preferences like those under discussion, it makes no difference whether she takes the long-term or the short-term perspective when evaluating the individual gamble; she will turn it down either way. Hence, contrary to a worry recently raised by Thoma (2019), we can predict the choices of a minimally rational person with the preferences under discussion even if we do not know which of these perspectives the person takes.

My argument relies on the assumption that one can only *determinately* choose for one's current self; for instance, gambles that a person will be offered in the future are not, strictly speaking, options for her today. I shall argue that it follows from this assumption (which I spell out in section 3), that a person with the combined attitude under discussion at some points in time faces a type of *Prisoner's Dilemma* between different "time-slices" of herself. And as in the *interpersonal* version of multiplayer Prisoner's Dilemma, there is, in the absence of an external constraint, only one rationalisable outcome of an *intrapersonal* version of the game, namely, a

---

individual preference is (ir)rational *per se*, even though some combinations of attitudes and certain combinations of attitudes and actions are (ir)rational. (I am grateful to Alexander Heape for making me see the need to add this clarification.)

“tragedy of the commons” (Lloyd 1833, Hardin 1968) where each time-slice turns down the individual gamble with which it is faced.<sup>8</sup>

I end the paper with some brief remarks on my most important assumption about rational sequential choice, that is, my assumption that rational people are not what has been called “resolute” choosers. In particular, I argue that if one takes seriously the game-theoretic analysis of the puzzle under discussion, then that provides at least some reason to assume that rational people with the preference structure in question are not resolute.

## 2 Samuelson’s colleague

The problem of interest is how risk averse agents should choose, when offered a gamble (i.e., risky prospect) with a positive objective (e.g., monetary) expectation that they have a preference against when considered in isolation, but that they would accept if offered as part of the collection of gambles that they predict they will be faced with in the future. To keep things simple, I will model this problem with help of the example that Paul Samuelson (1963) used when stating his “fallacy of large numbers”. Samuelson told a story of a “distinguished” colleague of his, who turned down an offer of a single 50-50 gamble between gaining \$200 and losing \$100, while confessing his willingness to accept a bundle consisting of 100 such (independent) gambles—which would indeed be a very attractive offer, carrying only about a 0.04% risk of losing money and having an expectation of \$5,000. Since slightly smaller bundles will be very attractive too, I will assume that there is some  $n < 100$  such that the colleague would prefer to enlarge the size of the bundle once the number of gambles in the bundle have reached  $n$ .

As Samuelson proved, the preferences of his colleague—whom I shall henceforth refer to as SC—cannot be represented as maximising expected utility, as traditionally defined (see fn. 1).<sup>9</sup> To prove SC’s inconsistency

---

<sup>8</sup>Buchak (2013: 222) briefly points out that the puzzle that I shall discuss bears some similarities to Prisoner’s Dilemma, without analysing the similarity further.

<sup>9</sup>Moreover, as Stefánsson (2018) shows, building on work by Tversky and Bar-Hillel (1983), SC’s combined attitude violates either Acyclicity—which states that if  $A$  is preferred to  $B$  and  $B$  is preferred to  $C$  then  $C$  is not preferred to  $A$ —or Event-Wise Dominance—

with expected utility theory, Samuelson made an assumption that I shall make too, namely, that SC would turn down the single gamble even if his wealth increased or decreased by the maximum that one can gain (\$19,800) or lose (\$9,900) from the first 99 gambles in the bundle. More generally, it will be assumed that the possible outcomes of the bundles that we are considering will not reverse the preferences towards the corresponding individual gambles.

By the reasoning explained in the introductory section, we seem to have the following argument that SC should take the long-term perspective. Suppose that SC gets the 100 gambles offered sequentially, that he knows this, and that he moreover learns the outcome of the previous gamble before being offered the next gamble.<sup>10</sup> Then if he at each choice point acts according to the preference he has at that point, he will turn down every gamble. However, if SC evaluates each gamble in light of the gambles that he predicts he will be later faced with, then he might ask himself whether he prefers, in the long-run, to acquire a smaller or a larger set of gambles, rather than asking himself whether he prefers to accept or reject the gamble on offer at each point in time. In that case, it might seem that he would accept each gamble, which he prefers to rejecting all gambles. Hence, it might seem that SC *should* frame the decision in terms of the long-run, since that way he acquires a bundle that he (*ex ante*) prefers to the “empty bundle” that he ends up choosing with the short-term framing.

So, the puzzle that is the topic of this paper can be formulated as follows. It would seem that instrumentally rational decision-makers, that is, decision-makers who make the choices that they expect to best satisfy their interests, will choose to formulate each decision-problem in terms of the long-run rather than in terms of the short-run. That seems to imply that an instrumentally rational decision-maker will *not* turn down an individual gamble if she would accept it as part of the collection of gambles that she

---

which implies that if there is a partition of the state space such that one gamble is (weakly) preferred to another gamble conditional on every event in the partition, then the first gamble should be (weakly) preferred to the second.

<sup>10</sup>If SC does not learn the outcome of the last gamble in a sequence before being offered the next gamble, then the decision he is faced with is a choice between a smaller and a larger bundle of gambles, as I shall interpret it, rather than a choice between accepting and rejecting an individual gamble.

predicts she will be faced with in the long-run. But how is that consistent with SC's preferences—that is, a preference against the single gamble but in favour of a bundle of 100 such gambles—being rationally permissible?

### 3 The impossibility of choosing for others

Let us look closer at the question of how SC should approach the decision about the 50-50 gamble between winning \$200 and losing \$100 that he is being offered *now*, assuming that he believes that he will be offered many more bets like this in the future, and assuming that he prefers accepting all of these gambles to accepting none of them.

As previously discussed, it is natural to think that to be instrumentally rational, SC must include anything that matters to the evaluation of the gamble when deliberating about whether to accept it or not. In particular, he should include information about the gambles that he predicts he will be offered. Let's accept this claim, at least for the sake of the argument. So, SC is required to take the "long-term perspective", when evaluating the gamble that he is now faced with.

How would taking the long-term perspective, rather than the short-term perspective, affect SC's current decision, that is, the decision between accepting and rejecting a single 50-50 gamble between winning \$200 and losing \$100? In the last section, we saw an argument that may have suggested that this difference in perspective makes SC accept rather than reject the gamble with which he is now faced. However, in the next section I shall argue that, perhaps surprisingly, he will, if rational, turn down the single gamble irrespective of whether he takes the long-term or the short-term perspective. But since there is nothing special about the gamble that he is *currently* faced with, it follows that he will turn down all such gambles that he will be offered in the future too. So, even if SC takes the long-term perspective, he will turn down all the gambles, even though he prefers the bundle of 100 gambles to no gamble at all.

The argument in the next section relies on the following pair of conceptually related by logically independent assumptions, that I shall spell out in the remainder of this section. First, the fact that a person, at a time

previous to  $t$ , intends to make a particular choice  $C$  when it comes to time  $t$ , does not make it impossible for the person to *not* make choice  $C$  when it comes to time  $t$ .<sup>11</sup> In other words, intending to perform a particular act at a later time  $t$  is consistent with *not* performing that act at time  $t$ . Second, to *determinately* choose, as we might put it, is something that one can only do for one's *current self*. One can intend for one's future self to make certain choices. And one can even prevent one's future self—or indeed someone else—from choosing an option, by making that option unavailable. But one cannot determinately choose for one's future self, nor for others. I leave open the possibility that one can *in some sense* choose for one's future self (and perhaps even for others); but that sense of choosing should then be kept distinct from what I call “determinate choosing”.

For instance, if I today tell a conference organiser that I would like the vegan dish at the conference dinner two weeks from now, then I am not determinately choosing the vegan option for my future self, given my terminology. Rather, I am choosing to limit the options of my future self in a particular way; for instance, by making it impossible for my future self to have a non-vegan dish without causing some trouble. In contrast, that my future self has the vegan dish is not an option for me today (cf. Hedden 2012). Similarly, when choosing between different types of coffee to bring your colleague from a cafe, you are not determinately choosing for the colleague, given my terminology. Rather, you are choosing how to limit your colleague's options.

Coming back to SC, it follows from the above way of understanding “to determinately choose”, that he cannot determinately choose now to accept some gamble that he will be offered in the future. He can at most determinately choose to accept or reject the gamble with which he is faced *now*.<sup>12</sup> Moreover, the fact that he now intends to accept some future gamble does not make it impossible for him to turn down the gamble when the

---

<sup>11</sup>Note that this assumption is consistent with Michael Bratman's (1987) influential account of intentions, according to which an intention to make a particular choice includes a *commitment* to make that choice. After all, even such a commitment is defeasible.

<sup>12</sup>How should we understand “now” and “current” (in “current self”)? That is a problem for the metaphysician, that I will not try to solve here. However, I contend that we have an intuitive notion of “now” and “current” that suffices to make plausible the claim that we can only determinately choose between options with which we are now faced, that is, options with which our current self is faced.



time arrives. And that is true even if he takes the long-term perspective, that is, even if he formulates his decision-problem by taking into account the fact that he predicts he will be offered multiple such gambles in the future. These future gambles are not options for SC today. Now, recall our above assumption that SC prefers to accept more rather than fewer of these future gambles. That means that SC now *hopes* that he will choose more rather than fewer of these future gambles. He might even now *intend* to accept these future gambles. But that does not mean that SC can now determinately *choose* more rather than fewer of these future gambles.

At the risk of being drawn into the muddy waters of the metaphysics of personal identity, we can formulate the above point by using the notion of a *time-slice* of a person. But my argument does not, I hope, depend on any particular stance on the metaphysics of personal identity. Nor does my argument depend on what Hedden (2015) calls *time-slice rationality*. That is, my argument is consistent with there being diachronic rationality requirements, as long as they do not at any time  $t$  trump the requirements of instrumental rationality that stem from what a person prefers at  $t$ . For instance, my argument is consistent with Bayesian conditionalization being a diachronic norm of belief change. My argument is of course also consistent with it being rationally permissible, and even consistent with it being rationally required, to consider one's life-time well-being when deliberating about a decision. But my argument is not consistent with, say, it being a general requirement of instrumental rationality that one makes some particular choice at time  $t$  whenever, and *because*, one hoped (or intended), at some point previous to  $t$ , that one would make that choice at time  $t$ .

Despite the above caveats, it should be evident that we can (metaphorically!) slice SC up however we like.<sup>13</sup> In particular, we can slice SC up into 100 time-slices, each slice corresponding to a gamble that he predicts he will be offered. Then we can formulate the observations from the preceding paragraphs as follows: Although SC's current time-slice prefers

---

<sup>13</sup>This might suggest that no (or hardly any) choices are ever available to us, since following through on any (or at least most) choice(s) takes some time. I believe that this boils down to the same issue as that raised in footnote 12, namely, the question of the maximum possible time-length of "current" in current self—which is an issue that I shall not try to resolve here.

to reject the gamble that it is offered, it nevertheless hopes that more rather than fewer of the future time-slices accept the gambles that they will be offered. And the same is true of all the other time-slices. However, each time-slice can only choose to accept or reject the gamble that it is offered.

## 4 An intrapersonal noncooperative game

Although each time-slice can only choose for itself, as I argued in the last section, each slice nevertheless should, when reasoning about the gamble with which it is faced, take into account the fact that other time-slices will be faced with gambles of the same type. After all, one should in general take the long-term perspective when formulating a decision-problem—or so we are at least assuming, for the sake of the argument—which implies, given the current formulation of SC’s decision-problem, that each time-slice should take into account the possible actions of other, future time-slices.

But how should each time-slice reason about its gamble given the gambles that other time-slices will be offered? Since each time-slice cannot determine how the other time-slices will act, each time-slice should ask itself what it would prefer to do given each of the possible combinations of actions available to future time-slices. In other words, each time-slice should frame its decision-problem as a *noncooperative game* with the other time-slices.

To analyse more formally SC’s decision-problem as a noncooperative game, let us start with some formal definitions and stipulations.<sup>14</sup> I shall continue to assume that SC will be offered 100 gambles sequentially and that we have divided SC into 100 corresponding time-slices; let’s now denote these  $SC_1, SC_2, \dots, SC_{100}$ . Let  $\prec^i$  be the *preference relation* of slice  $SC_i$ .<sup>15</sup> SC’s wealth before the start of the sequence of offered bets is denoted  $w$ ; before the final round of the sequence, SC’s wealth will be in the interval  $[w - \$9,900, w + \$19,800]$ ; let  $w^*$  be a random variable in this interval. Let  $\prec_{w^*}^i$  denote  $SC_i$ ’s preference *given* (or on the assumption that SC has) wealth

<sup>14</sup>I am particularly grateful to a referee and the handling editor for very useful suggestions that helped me clarify the following argument.

<sup>15</sup>For the present purposes, the domain of  $\prec^i$  can be very limited, only containing the decision to accept and the decision to reject the gamble with which  $SC_i$  is faced.

$w^*$ .  $A_i$  is the decision (by  $SC_i$ ) to accept gamble number  $i$ ;  $R_j$  the decision (by  $SC_j$ ) to reject gamble number  $j$ . Finally, let  $\mathcal{A}^n$  mean that  $n$  gambles will be accepted (but have not been resolved), and let  $\prec_{\mathcal{A}^n}^i$  stand for  $SC_i$ 's preference relation given that  $n$  future bets will be accepted.

We assume that the time-slices all have the same preference when it comes to any decision to accept or reject some particular gamble given any specified level of wealth, corresponding to the assumption that  $SC$ 's preference remains unchanged throughout the sequence. In particular, as the reader will recall, we have been assuming that for any  $SC_i$  and for any  $w^* \in [w - \$9,900, w + \$19,800]$ :

$$A_i \prec_{w^*}^i R_i \tag{1}$$

Less formally, each time-slices of  $SC$  prefers to turn down the gamble with which it is faced, when it considers that gamble only. We can think of the claim in (1) as formally capturing the assumption that  $SC$  prefers to turn down any individual gamble when he takes the *short-term perspective*.

However, we have also assumed that there is some  $n \in [1, 99]$ , such that for any  $i \in [1, 100]$  and for any  $w^* \in [w - \$9,900, w + \$19,800]$ :

$$R_i \prec_{w^*, \mathcal{A}^{m \geq n}}^i A_i \tag{2}$$

In other words, given that sufficiently many gambles will be accepted, each time-slice of  $SC$  prefers to accept the gamble with which it is faced (irrespective of  $SC$ 's wealth).

Now, for claim (1) to make any sense given claim (2), it would seem that there also has to be some  $n \in [0, 99]$  such that for any  $SC_i$  and for any  $w^* \in [w - \$9,900, w + \$19,800]$ :

$$A_i \prec_{w^*, \mathcal{A}^{m \leq n}}^i R_i \tag{3}$$

That is, there is some number  $n$  such that any time-slice will (irrespective of  $SC$ 's wealth) prefer to turn down the gamble with which it is faced given that no more than  $n$  future gambles will be accepted.

In particular, we have been assuming (as did Samuelson in his original

argument) that for any  $SC_i$  and for any  $w^* \in [w - \$9,900, w + \$19,800]$ :

$$A_i \prec_{w^*, \mathcal{A}^0}^i R_i \quad (4)$$

That is, given that no future gambles will be accepted, a time-slice prefers to turn down the gamble with which it is faced (irrespective of SC's wealth).

From the claim in (2) it follows that *if*  $SC_1$  is convinced that, say, all remaining 99 time-slices will accept—more generally, if  $SC_1$  is convinced that sufficiently many will accept—then  $SC_1$  will also prefer to accept. However, from (4) it follows that, no matter what the first 99 time-slices choose, and no matter the outcomes of any potential earlier bets,  $SC_{100}$  will prefer to reject. Assuming that  $SC_{99}$  foresees the preference of its successor, and believes that the successor will act on its preference (more on which in the next section),  $SC_{99}$  will also prefer to reject the gamble with which it is faced. But the same applies then to  $SC_{98}$ , and so on, all the way to  $SC_1$ .

In sum, we have seen that, although by (2), there are beliefs that  $SC_1$  in principle could hold such that it would prefer to accept, nevertheless by *backward-inductive* reasoning,  $SC_1$  will prefer to reject, and so will all other time-slices of SC.

We can now see that for the last time-slice, the noncooperative game in question has the payoff structure of a Prisoner's Dilemma. And the same is true of some other "late" time-slices.<sup>16</sup> In contrast, for the first time-slice, the noncooperative game in question does not have the payoff structure of a Prisoner's Dilemma. And the same is true of some other "early" time-slices. (It might be worth emphasising that which time-slices count as "late", and which count as "early"—that is, which time-slices face a game with the payoff structure of a Prisoner's Dilemma and which do not—is irrelevant for the conclusion, as long as the very last slice faces a game with this payoff structure.)

To illustrate the above two claims, let us consider the situation of a single time-slice under the simplifying assumption that the other time-

---

<sup>16</sup>I should note that the game itself is however somewhat different from standard formulations of multiplayer Prisoner's Dilemma. In particular, players in a Prisoner's Dilemma always act without (perfect) information about *simultaneous* action of some other player. The same is not true of the different time-slice of SC, since they always act one at a time.

slices all act in the same way, that is, we assume for now that the other time-slices either all accept or all reject the gambles that they are offered. The below table represents such a situation. The numbers in the cells are *ordinal* utilities, that is, they represent how the time-slice in question orders the bundles of gambles in terms of preferability, but they need not correspond to *how much* the time-slice prefers one bundle over another; the variables,  $n$  and  $m$ , denote ordinal utilities too.

	Others accept	Others reject
Accept	$m > n$	1
Reject	2	$n > 1$

Table: Each time-slice's decision-problem

The outcome represented by '2' is one where the time-slice whose decision-problem is being represented rejects the gamble with which it is faced while all the other 99 time-slices accept their gambles. The outcome represented by '1' is one where the time-slice in question accepts the gamble with which it faced while all the other 99 time-slices reject their gambles. The outcome represented by 'm' is one where the time-slice in question accepts the gamble with which it is faced and so do all the other 99 time-slices. Finally, the outcome represented by 'n' is one where the time-slice whose decision-problem is being represented rejects the gamble with which it is faced and so do all the other time-slices.

Note that the table leaves open the relationship between '2' and 'm'. In fact, the relationship is different depending on which time-slice we are considering. From the assumption formalised by (2), it follows that if the time-slice in question is  $SC_1$  (or perhaps some other "early" time-slice), then  $m > 2$ . That is, *given that* the other time slices *will* accept their gambles,  $SC_1$  prefers to do so too. In contrast, by the assumption formalised by (4), it follows that if the time slice in question is  $SC_{100}$ , then  $m < 2$ . For even if the earlier time-slices *have accepted* their gambles,  $SC_{100}$  prefers to reject. So, for  $SC_{100}$  (and perhaps some other "late" slices), the game has the payoff structure of a Prisoner's Dilemma.

Now, since  $SC_{100}$  prefers to reject the gamble even if the other slices accept, and assuming that  $SC_1$  reasons backward-inductively (an assumption we shall get back to in the next section) and moreover believes that all

slices act on their preferences,  $SC_1$  believes that it will only get to choose between '1' and 'n' (that is, it believes it will find itself in the 'Others reject' column). So, the relationship between '2' and 'm' ends up playing no role in  $SC_1$ 's reasoning. Thus,  $SC_1$  rejects the gamble with which it is faced. The same reasoning of course applies to  $SC_2$ ,  $SC_3$ , and so on; that is, they all reject the gambles with which they are faced. And this of course still holds if we complicate the decision-problem with which each slice is faced by including all the logically possible combinations of "Accept" and "Reject" by other time-slices (i.e., by removing the simplifying assumption that the "other" slices all act in the same way).

The implication of the above argument for the "time-extended" SC is that if he acts instrumentally rationally at each point in time—that is, if he always chooses in line with his preference—he will reject all the 100 gambles on offer. So, even though SC prefers accepting 100 gambles to accepting no gamble, he will, by acting rationally at each point in time, turn down every gamble.<sup>17</sup>

The above conclusion is reminiscent of the *tragedy of the commons* (Lloyd 1833), that is, the fact that individually rational players with a preference structure like "late" time-slices of SC will, by acting instrumentally rationally, together cause an outcome that is worse, from their own point of view, than an outcome that they (in theory) could ensure by acting together.<sup>18</sup> And just like the traditional *interpersonal* tragedy, the *intrapersonal* tragedy under discussion can be avoided by adopting some system of "mutual coercion" (Hardin 1968). Two ways of implementing such coercion seem particularly salient in the present context.

First, if SC could somehow turn his sequential choice problem into a static one, then he could secure himself all 100 gambles without ever

---

<sup>17</sup>By analogous reasoning, we get a similar conclusion for a person who has a preference in favour of the individual gamble but against the bundle of 100 gambles: Such a person will accept each gamble, even if she takes the long-term perspective. Such a combination of attitudes seems much less common than attitudes like SC's. Moreover, such attitudes also seem harder to rationalise than SC's. For instance, it would be hard to rationalise such attitudes by risk seekingness. After all, it is hard to understand why the risk seeking would reject the bundle of 100 gambles.

<sup>18</sup>Kavka (1991) discussed intrapersonal, but *synchronic*, versions of the Prisoner's Dilemma, and concluded (as I do) that avoiding the tragic outcome is no less difficult in the intrapersonal version of the game than in the traditional interpersonal game.

acting irrationally (i.e., counter-preferentially). So, SC should be willing to pay some price for turning the problem into one of choosing *in one go* between a bundle consisting of 0 gambles and a bundle consisting of 100 gambles. Second, if SC could somehow remove the option to turn down the gambles that he will later be offered, he would rationally choose to do so. For instance, SC should be willing to pay someone to limit his options in this way—much like Ulysses in Homer’s *Odyssey* ordered his crew to tie him to the mast, thereby removing the option of leaving the ship.

However, if SC is really faced with the sequential choice problem, and assuming that he is instrumentally rational, he will turn down each gamble if he has that option. So, just like rational people in a multiperson Prisoner’s Dilemma, a rational SC faces a sure tragedy: The tragedy of the risk averse.

## 5 Should SC be resolute?

Before concluding this paper, I shall briefly address the question of why, unlike some defenders of alternatives to expected utility theory, I do not assume that a rational SC is what is called a *resolute* chooser (McClennen 1990). To put it informally, a chooser is resolute, in this sense, if she decides at the start of some sequential choice problem to implement a particular plan, and chooses accordingly at every choice point, *even if that means choosing against her preference at some choice point(s)*. In the case of SC’s sequential choice problem, the resolute choice could be to, for instance, decide at the start of the sequence to accept all gambles, even though that would require choosing counter-preferentially at some choice points.

Now, not being resolute makes decision-makers like SC—more generally, decision-makers who are not expected utility maximisers—vulnerable to certain costs in sequential decision-problems. We have already seen instances of this: If possible, SC will choose to pay a price to remove the option of rejecting the gambles in the sequence. He will also willingly pay a price for changing his sequential decision-problem into a one-off (i.e., static) decision. And note that this willingness to pay to affect the structure of the decision-problem is not due to predicted weakness of the will or preference change.



SC thus differs from any expected utility maximiser, who will end up with the same outcome irrespective of whether the decision-problem is static or sequential (as long as the possible outcomes and the relevant probabilities are the same), unless she suffers from weakness of will of some sort or changes her mind during the sequence. Hence, such an expected utility maximiser will neither pay a price to bind herself nor will she pay to turn a sequential choice problem into a one-off decision. But neither would a resolute SC, or so the argument goes, since he would simply resolve to accept each gamble on offer, despite his preference.

For the above reasons, many defenders of alternatives to expected utility theory—that is, people who think that preferences that violate the axioms of the theory can nevertheless be rational—have been tempted by the view that rational people who are not expected utility maximisers are resolute choosers (see e.g. McClennen 1990, Buchak 2013, Thoma 2017, 2019). I think that is a mistake: It can never be rational, at time  $t$ , to choose against what one truly prefers, at time  $t$ ; hence, a rational person will not be a resolute chooser.<sup>19</sup> In fact, I think that those who defend alternatives to expected utility theory should simply accept that certain preferences are such that an agent with those preferences can find herself in a situation where she would get an outcome that is better, from her own point of view, had she had different preferences. But that by itself does not give the person a reason to change her preferences, assuming for instance that the situation in question is sufficiently rare or unlikely to occur.<sup>20</sup>

That resolute choice is generally irrational is however not a claim that I can hope to defend in this paper. Instead, I will simply point out that those who think that SC should choose resolutely will, if the argument from the last section is sound, have to accept a pretty radical implication.

As the reader will recall, I argued in the last section that SC faces a noncooperative game. Granted, SC faces a game between different time-

---

<sup>19</sup>This is not to say that rational people cannot be what Wlodek Rabinowicz (1995) calls *wise* choosers. That is, rational people may take into consideration *past* choices and intentions, etc., in the sense of making them part of the description of their *current* options. But if such considerations alter a person's attitude to a gamble, then that person does not have a simple preference structure like SC.

<sup>20</sup>So, in my view, a person with cyclical preferences should not change her preferences *because* she could otherwise (in theory) be money pumped.



slices of himself, rather than with other people; and one's relationship to different time-slices of oneself may, in important ways, differ from one's relationship to other people. However, the difference in question, whatever it may be, is irrelevant to the argument of this paper. Even though SC's relationship to other people differs in many respects from the relationship between different time-slice of SC, the two relationships share an important feature: One time-slice of SC cannot determinately choose for another time-slice of SC, no more than SC can determinately choose for another person. Nor can one time-slice count on some external force that ensures that another time-slice "cooperates". And that suffices to make the game that each time-slices has to play a noncooperative game.

In light of the above, it is hard to see how it could be (instrumentally) rational for SC to choose resolutely, unless it can be (instrumentally) rational for different players in an *interpersonal* noncooperative game, such as the Prisoner's Dilemma, to choose resolutely. So, those who claim that SC should be resolute, will, if my argument is sound, have to accept that people facing some noncooperative games should be resolute. Occasional scholars may be willing to accept that implication (in particular, McClenen 1997 and Gauthier 1996). But I think that most of those who feel tempted by the idea that those who are not expected utility maximisers should be resolute in *non-strategic* sequential choice problems, will nevertheless find it hard to accept that players in, say, a Prisoner's Dilemma should be resolute.

To conclude, there may be reasons to regret being risk averse. Most obviously, the risk averse should expect to do worse in the long-run than the risk neutral. However, doing worse than one would expect to do, had one been different, does not make one irrational. In fact, the fate of the risk averse is tragic precisely because by acting instrumentally rationally they choose prospects that are worse, by their own lights *ex ante*, than the prospects that they could have chosen.

## References

- Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Center for the Study of Language and Information.
- Buchak, L. (2013). *Risk and Rationality*. Oxford University Press.
- Gauthier, D. (1996). Commitment and choice. In Farina, F., Vannucci, S., and Hahn, F., editors, *Ethics, Rationality, and Economic Behaviour*, pages 217–243. Oxford University Press.
- Hardin, G. (1968). The tragedy of the commons. *Science*, 162(3859):1243–1248.
- Hedden, B. (2012). Options and the subjective ought. *Philosophical Studies*, 158(2):343–360.
- Hedden, B. (2015). Time-slice rationality. *Mind*, 124(494):449–491.
- Kavka, G. S. (1991). Is individual choice less problematic than collective choice? *Economics and Philosophy*, 7(2):143–165.
- Lloyd, W. F. (1833). *Two lectures on the checks to population*. Oxford University Press.
- McClennen, E. F. (1990). *Rationality and Dynamic Choice: Foundational Explorations*. Cambridge University Press.
- McClennen, E. F. (1997). Pragmatic rationality and rules. *Philosophy and Public Affairs*, 26(3):210–258.
- Rabinowicz, W. (1995). To have one's cake and eat it, too: Sequential choice and expected-utility violations. *Journal of Philosophy*, 92(11):586–620.
- Samuelson, P. A. (1963). Risk and uncertainty: A fallacy of large numbers. *Scientia*, (98):108–113.
- Stefánsson, H. O. (2018). Is risk aversion irrational? *Synthese*, <https://doi.org/10.1007/s11229-018-01929-5>.

- Thaler, R. H., Tversky, A., Kahneman, D., and Schwartz, A. (1997). The effect of myopia and loss aversion on risk taking: An experimental test. *The Quarterly Journal of Economics*, 112(2):647–661.
- Thoma, J. (2017). *Advice for the Steady: Decision Theory and the Requirements of Instrumental Rationality*. PhD thesis, University of Toronto.
- Thoma, J. (2019). Risk aversion and the long run. *Ethics*, 129(2):230–253.
- Tversky, A. and Bar-Hillel, M. (1983). Risk: The long and the short. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(4):713–717.