Preprints of the
Max Planck Institute
for Research on Collective Goods
Bonn
2005/17

Why humans care about sunk costs while animals don't. An evolutionary explanation

Felix Höffler

Preprints of the
Max Planck Institute
for Research on Collective Goods

Bonn 2005/17

Why humans care about sunk costs while animals don't. An evolutionary explanation

Felix Höffler

September 2005

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
http://www.mpp-rdg.mpg.de

Why humans care about sunk costs while animals don't. An evolutionary explanation

Felix Höffler*

September 20, 2005.

Abstract

While humans often care about sunk investment, animals are not subject to this sort of sunk cost behavior or "Concorde fallacy". This paper investigates a simple two stage decision problem under uncertainty. At the second stage, subjects can commit the Concorde fallacy by sticking to the first stage decision, independent of the state of nature revealed in-between. We investigate whether this can be beneficial in a standard payoff monotonic adaptation process. Committing the Concorde fallacy reduces the payoffs but accelerates the adaptation since it acts like "self-punishment". It will, however, not only reduce the population growth rate in the long run but also the population size at any point in time in a biological evolutionary process. In this sense, animals can never benefit from the Concorde fallacy. By contrast, risk averse humans using a simple learning rule can increase their expected utility by committing the Concorde fallacy. Risk aversion gives an extra benefit to a behavior that more rapidly learns to avoid bad outcomes. If the wrong initial decision leads occasionally, albeit very infrequently, to a very low payoff, then risk averse humans will be better off by committing the Concorde fallacy.

JEL-Classification: C73, D81, D83

Keywords: Concorde fallacy, Sunk costs, Evolutionary game theory, Replicator Dynamics, Risk aversion

*Max Planck Institute for Research on Collective Goods. Kurt-Schumacher-Str. 10, 53113 Bonn, Germany. ++49(0)228 9141646. hoeffler@coll.mpg.de. I would like to thank Tilman Börgers, Werner Güth and Georg Nöldeke for stimulating discussions. Francis Ratnieks introduced me to the biological literature on the Concorde fallacy. Angelika Praus provided excellent research assistance. All remaining errors are mine.

1 Introduction

The "Concorde fallacy" and "sunk cost effects" refer to the same problem of suboptimally attributing attention to past investments while only future, marginal effects should be taken into account. While the first term has been used by biologists in the context of animal behavior, the second term is familiar to economists and psychologists alike, referring to human behavior.

Everyday experience tells us that humans are vulnerable to sunk cost behavior: When we buy a ticket for the opera and – on the evening of the performance – recognize that there is an interesting football match on TV, which we would actually prefer to watch, we feel somehow obliged to go to the opera in order to avoid the sense that we have "wasted" the money on the opera ticket. This sort of behavior is well documented in many psychological experiments.

Sunk cost behavior also plays a crucial role in many economic contexts, in particular with respect to follow-up investment decisions. The term "Concorde fallacy" refers to the project of building the "Concorde", the first civil supersonic aircraft. It became obvious quite early that it would not be a commercial success (Woolley (1972), Edwards (1973)). However, a strong argument to finalize the project was that otherwise the sunk investment would have been in vain.

Not only investment decisions are subject to sunk cost behavior. "Full-cost pricing" can represent another form of the Concorde fallacy. An interesting example is reported by Thomas (1981), 87. Thomas A. Edison first had only limited financial success with his famous invention, the electrical lamp. Bulbs were not selling well and capacities were underutilized. Edison proposed to fully use the capacities and to sell the additional lamps below fully allocated cost but above marginal cost. His managerial experts strongly opposed this proposal of "selling below cost" as unprofitable.[1]

In contrast to humans, animals seem less prone to such fallacies. Biologists' have investigated many species and have not found a clear incidence where animals commit the Concorde fallacy. This has given rise to the question: "Are humans less rational than lower animals?" the title of a review article by Arkes and Ayton (1999) comparing experimental evidence for humans and animals.

The current paper tries to provide an evolutionary interpretation of why animals are less likely to care about sunk cost than humans. We take up the psychological idea of "overgeneralization": Humans apply a certain rule (Don't waste!) which often leads to optimal decisions, sometimes in situations where this is not adequate. Caring for sunk cost can be beneficial if it helps to get important decisions right at the cost of errors in less important situations.

This idea is formalized using a simple two stage decision problem. At stage one an individual has to chose between two initial alternatives. At stage two she can either carry on with the initial decision or switch to a different alternative. Which of the two is optimal depends on the state of the world which is revealed between stage one and stage two. The Concorde fallacy consists of carrying on

[1] The importance of accounting profits provides another example that we care about sunk cost. Since accounting profits are determined to a high degree by depreciations, they reflect sunk investments.

independent of the state of the world. The analysis focuses on the simplest situation, in which one of the two alternative always yields higher payoffs, provided an optimal stage two decision.

We investigate how the right initial decision is "learned" in an adaptation process. Three different processes are considered: (i) A standard population game of biological evolution; (ii) individual decision-making of a boundedly rational risk averse agent using a reinforcement learning rule; and (iii) imitation learning in a large population. All three processes can be discussed using a similar replicator dynamics formulation.

The key effect is that committing the Concorde fallacy can serve as a self-punishment mechanism for wrong initial decisions: this can speed up the adaptation process. Thus, there is a general trade-off between accelerating the adaptation (getting the initial decision right) and the average payoff in each of the underlying decision problems (lower payoffs whenever one sticks to the initial decision although this is not optimal). For linear objective functions (e.g. the number of offsprings in biological evolution, case (i)), the latter effect always dominates. Therefore, the results are in line with the observations that animals are not subject to sunk cost behavior. For concave objective functions (e.g. utility of a risk averse individual, cases (ii) and (iii)), committing the Concorde fallacy is always beneficial if only the worst of all possible outcome is sufficiently bad. Risk aversion gives a particular weight to avoiding bad outcomes, even if they are very infrequent. Committing the Concorde fallacy can help here by smoothing payoffs. It reduces payoffs on average but accelerates learning such that drastic mistakes become rare more rapidly.

The remainder of the paper is organized as follows. Section two provides an overview on the empirical results on the Concorde fallacy. Section three introduces the decision problem. Section four discusses the evolutionary applications of maximization of fitness and learning and derives the replicator dynamics. Section five contains the major results which are discussed in the last section.

2 Empirical evidence for Concorde fallacy and sunk cost effects

There is extensive empirical biological research on the "Concorde fallacy". Though it impossible to prove the non-existence of an empirical phenomenon like the Concorde fallacy, it is fair to state that it is highly unlikely that animals commit the Concorde fallacy. There is at least no clear cut example in which animal behavior could not equally well be explained in terms of "adaptive" behavior, i.e. in line with traditional explanations of maximizing fitness (see Curio (1987) and Arkes and Ayton (1999) for summaries of the experimental literature). A good example for this stems from the early discussion about the question: "Do digger wasps commit the Concorde fallacy?" the title of a paper by Dawkins and Brockman (1980). Female digger wasps lay a single egg in a burrow, and

before doing so provide the nest with food (paralyzed katydids). Occasionally, a second wasp moves in and the two jointly provide for the nest. When the two meet at the nest, they fight for the nest. Dawkins and Brockman report that a wasp will fight more intensely the more katydids it has already stored in the nest. This could accord with a Concorde fallacy, since the wasp should care for the total number of katydids, not only the "own" katydids. Dawkins and Brockman, however, convincingly argue that this behavior need not be "maladaptive" if the wasps cannot count the number of katydids. Then their own effort of supplying the nest is the best estimator available to the wasps for the total number of katydids in the nest. It is in this sense "rational" to condition the intensity of the fight only on their own past investment in the nest.

Empirical results are different for humans. For humans, the "sunk cost" behavior is a phenomenon well-established by many experiments. Arkes and Ayton (1999) provide a summary of the literature and report a typical experiment: Season tickets for the Ohio University Theater were sold at three different price levels: $15, $13, and $8. The allocation to the different price classes happened randomly, thus, there is no correlation between the willingness to attend the plays and the price levels to be expected. However, those with a higher price attended the plays more often. One explanation is the sunk cost effect: because they had invested more in the past, they felt more obliged to use the tickets.

There exist at least three main explanations in the psychological literature for the "Concorde fallacy" among humans. First, using the framework of *prospect theory* (Kahneman and Tversky (1979)), Garland and Newport (1991) argue that human decision-makers take the endowment before making the initial investment as the reference point for their decision. Abandoning a project in which one has already invested in the past creates a loss with certainty. Decision-makers "gamble for resurrection" because they find themselves at the convex part of the utility function, which makes them risk loving, implying that they are willing to undertake follow-up decisions with a negative expected value (see also Arkes and Blumer (1985) and Zeelenberg and van Dijk (1997)).

Second, *reputational effects* are often mentioned as a reason for sunk cost behavior. Abandoning an unsuccessful project that one has started makes it obvious that the initial decision was wrong. Carrying on avoids (at least for some time) disclosing one's own mistake. This would explain well why decision makers are more likely to correct a past decision which had been taken by others than to correct own decisions. McCarthy, Schoorman, and Cooper (1993) provide evidence of such behavior for the decision to expand an existing business. Someone who has created the business is more likely to expand it than someone who has acquired an existing business. Arkes and Ayton (1999), 592, cite further evidence of reputational effects.[2]

[2] Several papers from economic theory also investigate reputational effects of sunk costs. Boot (1992) maintaines that post-takeover divestitures are often value increasing because of the reluctance of the old management to alter (wrong) past investment decisions. Managers reveal that they are of the low ability type by correcting past decisions. In this paper, bad managers are less prone to correct wrong decisions. Kanodia, Bushman, and Dichhaut (1989) also rationalize sunk cost behavior by career concerns of managers who do not want to be

Third, the *overgeneralization of rules* plays an important role (Arkes and Ayton, 1999, 592). "Don't waste" is an important rule humans have to learn. Children shall eat up their meals, students should finish their studies and so on. This is, however, sometimes not optimal. Therefore, the application of a rule, which originally was meant to prevent the waste of resources, occasionally leads to suboptimal decisions because decision-makers cannot determine when the application of the rule is adequate and when not. Arkes and Ayton argue that this is an important reason why humans commit the Concorde fallacy. They also cite evidence that the only experimental subjects reported to behave more rationally are children below the age of 6 (Arkes and Ayton, 1999, 598) – thus, it seems that we teach our children to become irrational and to commit the Concorde fallacy.

Although all three (and perhaps even more) psychological explanations are probably important, the arguments presented here are most closely related to the learning argument in the overgeneralization hypothesis.

3 A sunk cost problem

Sunk cost problems occur in dynamic decision problems only. Figure 1 depicts the minimum structure of such a problem.

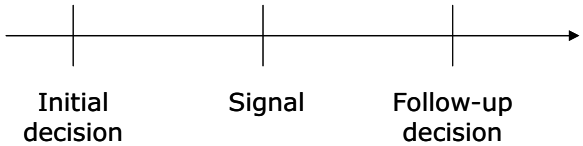

Figure 1: Time Structure of a Sunk Cost Problem

There are at least two consecutive, interrelated decision problems. The initial decision involves options that in some sense are associated with options of the follow-up decision. One typical example is that the initial decision is a project choice, while the second regards whether to continue or expand the project or to abandon it. In-between the initial decision and the follow-up decision, there is some payoff-relevant signal about the state of the world. An optimal follow-up decision will be conditioned upon this signal. Committing the Concorde fallacy means ignoring the signal and selecting the "associated option" in any event.[3]

regarded as low ability types. While in these papers there is no link between the manager's and the owner's interest, Sliwka (2005) considers the case where the manager also cares about the firm's payoff. He finds that it is the good manager who is less willing to change past decisions and argues that therefore it might be that the owner is better off with a bad manager (who is willing to correct his decision) than with a good but stubborn manager.

[3] Game theory has some difficulties in grasping this dimension of the Concorde fallacy. When confronted with a sunk cost problem, most people can identify what the "associated" action is, e.g. carrying on with a project. In an extensive form game, however, the Concorde fallacy has only one property: it yields suboptimal payoffs. However, there may by many other suboptimal options which we would not consider as committing the Concorde fallacy. In the game theoretic sense, "associated" is just nomenclature.

We want to analyze a simple decision problem with two initial alternatives, interim information and the possibility of committing the Concorde fallacy at stage two. We make the extreme assumption that one of the two initial alternatives is better, independent of the state of the world (we discuss this assumption in the conclusions). Figure 2 shows the decision problem. For illustrative purposes, consider the following situation. You want to climb a mountain. There are two possible ways, the right one (I_A) and the wrong one (I_B). In the morning you have to take one of the two alternatives. The success of the tour depends on how the weather develops during the day: the weather can be good (θ_A) or bad (θ_B). If you take the right way I_A and you have fair weather you arrive safely at the summit, you enjoy the trip and receive a payoff of A. If you have taken the right way but the weather becomes bad, it is best to decide at stage two to return (i_B), which yields $C < A$. Alternatively, you could commit the Concorde fallacy and think: I have been walking for so long, thus I want to carry on (i_A) to get to the summit anyway - this will be very exhausting and unpleasant, yielding $D < C$. If you have chosen the wrong way (I_B) and the weather is good, you will have a similar choice: you can carry on (i_B) although you will have realized that you are on the more difficult path. You will get to the summit, but again it will be an unpleasant and exhausting journey, yielding D. It is, however, better to return (i_A), which yields $C > D$. If you have opted for the wrong way and the weather is bad, you have no chance to get to the summit nor to return but you have to stay in the mountains overnight, which is very unpleasant and dangerous, yielding $B < C$.

In this decision problem the options I_A (I_B) and i_A (i_B) shall be regarded as "associated" (that is "carrying on" in the example). The "good" state of the world θ_A (good weather) occurs with probability α, and θ_B (bad weather) occurs with probability $(1-\alpha)$. As discussed, the state of the world is observed before making the follow-up decision.

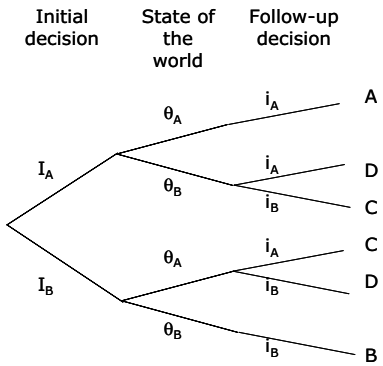

Figure 2: Payoff Structure

Throughout we assume:

$$C > D, \tag{1}$$
$$A > C > B. \tag{2}$$

Assumption (1) implies that it is sometimes not optimal to take the "associated" decision. If the decision-maker has chosen I_A (I_B), and the "other" state of the world θ_B (θ_A) occurs, it is better to "switch" to i_B (i_A). Choosing the associated action i_A (i_B) is interpreted as committing the Concorde fallacy.

Assumption (2) implies that I_A is the optimal stage one decision, provided an optimal stage two decision. Assuming an optimal follow-up decision, the payoffs can be represented in a reduced payoff matrix:

Table 1: Reduced form payoff matrix

	I_A	I_B
θ_A	A	C
θ_B	C	B

Using this representation, investigating the Concorde fallacy is equivalent to asking: Do well-defined environments exist in which reducing C (to some $D = C - \varepsilon$) can be beneficial? The following section defines three such environments.

4 Adaptation Processes

Humans seem to behave differently from lower animals with respect to the sunk cost problem. The assumption of this paper is that this is the result of different adaptation processes. For humans, we believe that behavior is the result of (social) learning. This is in line with the observations that children seem to be less prone to commit the Concorde fallacy compared to grown-ups. For (lower) animals, we assume that behavior is genetically determined by a process of mutations and selection.

Both processes exhibit many differences with respect to the adaptation mechanics. While evolution needs thousands or millions of generations, learning can happen in a far shorter time frame. The focus, however, of this paper is not on explaining different outcomes by differences in the mechanics of the process (although this also might play an important role). Of the two ingredients of the model, the adaptation mechanics and the objective function, we hold the first constant when comparing humans to lower animals in order to be able to identify the effect of altering the latter.

In doing this we can exploit the fact that evolution and (social) learning can be represented using the same formalism, namely the replicator dynamics. Therefore, we will discuss three different adaptation processes (one for animals, two for humans) for the decision problem of Figure 2, which can all be represented using the replicator dynamics. As a starting point, in each case we assume that the second stage decision is taken optimally in each context. The adaptation process therefore refers to the first stage only. Building on these

results, we will answer the key question of the paper: i.e. whether deviating from behaving optimally in the second stage, namely earning some $D < C$, can have positive effects measured by the objective function.

Biological evolution Consider a population of animals where each animal is programmed to either play I_A or I_B. The population size is denoted by N and is large but finite. Each animal inherits its strategy from its parents. The payoff equals the number of offspring in each period. Often we will refer to the payoff from playing I_A as $x = \alpha A + (1-\alpha) C$, and to the payoff from playing I_B as $y = \alpha C + (1-\alpha) B$. Call $p(t)$ the fraction of animals in the population which choose I_A. The fraction of I_A-types in $t+1$ will be equal to:

$$p(t+1) = \frac{p(t)(1+x)}{p(t)(1+x) + (1-p(t))(1+y)}. \tag{3}$$

The total number of offspring at time t is given by:

$$N(t) \cdot [p(t)x + (1-p(t))y]. \tag{4}$$

Using standard techniques (see Weibull (1995), 124, and Samuelson (1997), 65) the discrete dynamics can be approximated by a continuous time formulation in which the growth rate of the population is given by:[4]

$$\frac{dN_t/dt}{N_t} = \pi(t) = p(t)x + (1-p(t))y, \tag{5}$$

and the population share $p(t)$ can be approximated by the following expression

[4] We have in mind an overlapping generations model in which, in each instant in time, reproduction and death takes place for a fraction r of animals per unit of time, where a unit of time is $\tau = 1/r$. For details, see e.g. (Weibull, 1995, 124). Thus:

$$\begin{aligned} \frac{dN}{dt} &= \lim_{\tau \to 0} \frac{N_{t+\tau} - N_t}{\tau} = \lim_{\tau \to 0} \frac{N_t + \tau p x + \tau (1-p) y - N_t}{\tau} \\ &= N_t (px + (1-p) y), \end{aligned}$$

implying

$$\frac{dN_t/dt}{N_t} = (px + (1-p) y).$$

in which x and y have been substituted by the underlying payoffs:[5]

$$\begin{aligned}\frac{dp}{dt} &= p(t)(1-p(t))x - p(t)(1-p(t))y \\ &= p(t)(1-p(t))(x-y) \\ &= p(t)(1-p(t))[\alpha(A-C)+(1-\alpha)(C-B)]. \end{aligned} \tag{6}$$

Equation (6) reflects the standard replicator dynamics.

Social (imitation) learning Consider a large population N of bounded rational players who play pure strategies. After each game, each player observes the payoff and strategy of another randomly selected player. He will imitate this player with probability π/A, where π now denotes the observed payoff.[6] The fraction $p(t)$ of players playing strategy I_A at time t is given by a difference equation similar to (3), except for a normalization term $1/A$:

$$\begin{aligned} p(t+1) &= p(t) + p(t)(1-p(t))\frac{\alpha A + (1-\alpha)C}{A} \\ &\quad - (1-p(t))p(t)\frac{\alpha C + (1-\alpha)B}{A}, \end{aligned} \tag{7}$$

which can be approximated by a continuous time formulation:

$$\frac{dp}{dt} = p(t)(1-p(t))\frac{\alpha(A-C)+(1-\alpha)(C-B)}{A}. \tag{8}$$

Individual (reinforcement) learning Consider a bounded rational player who repeatedly plays the game. She employs a simple learning rule which is made conditional on her own past payoff. She uses a mixed strategy, where $p(t)$ denotes the probability to select I_A at time t. If she played I_A and received a payoff π, then she would reduce the deviation from selecting I_A with certainty, which is given by $1-p(t)$, by the factor π/A. For I_B, vice versa. This payoff monotonic learning rules yields the following as an expected probability for playing I_A at time $t+1$:

$$\begin{aligned} p(t+1) &= p(t) + p(t)(1-p(t))\frac{\alpha A + (1-\alpha)C}{A} \\ &\quad - (1-p(t))p(t)\frac{\alpha C + (1-\alpha)B}{A}. \end{aligned}$$

---

[5]

$$\begin{aligned} p(t+1) &= \frac{p(t)+p(t)x}{p(t)(1+x)+(1-p(t))(1+y)} \\ &= \frac{p(t)+p(t)x}{1+p(t)x+(1-p(t))y} \\ p(t+\tau) &= \frac{p(t)+\tau p(t)x}{1+\tau p(t)x+\tau(1-p(t))y} \\ \lim_{\tau\to 0}\frac{p(t+\tau)-p(t)}{\tau} &= \lim_{\tau\to 0}\frac{1}{\tau}\left(\frac{p(t)+\tau p(t)x}{1+\tau p(t)x+\tau(1-p(t))y} - p(t)\right) \\ &= (6) \end{aligned}$$

[6] This is a payoff monotonic learning rule, similar, though not identical, to a "proportionate observation rule", discussed by Schlag (1998).

When analyzing the payoffs in expected terms, both learning processes can thus be described by the same formalism, namely the replicator dynamics (8).

The solution to the ordinary differential equation (6) describing the biological evolutionary process is given by:

$$p(t) = \frac{e^{\widetilde{z}t}}{e^{\widetilde{z}t} + A_0}, \tag{9}$$

where $\widetilde{z} = \alpha(A - C) + (1 - \alpha)(C - B)$ and A_0 reflects the initial condition:

$$p(0) = \frac{1}{A_0 + 1}. \tag{10}$$

For the learning process, the same solutions apply, except for the fact that $\widetilde{z}$ has to be replaced by $z = \frac{\alpha(A-C)+(1-\alpha)(C-B)}{A}$.

The formal structure of all three evolutionary processes is similar. Therefore, we can state some general results which apply to all of them. The equation of motion (6) has two obvious characteristics: First, $dp/dt > 0$, since $(x - y) > 0$. All three evolutionary processes will drive out the suboptimal choice of I_B in the long run, since I_B unambiguously yields a lower payoff.

Lemma 1 *$p \to 1$ for $t \to \infty$ for all evolutionary processes described above.*

A second observation concerns the "switching payoff": that is, the payoff that results when adapting to the state of the world implies switching to a different follow-up decision. Its effect on the learning process depends upon the probability of the good state of the world, α. The payoff difference between the initial optimal and the suboptimal initial decision, which is given by the factor $[\alpha(A - C) + (1 - \alpha)(C - B)]$ in (6), increase in C only if $\alpha > \frac{1}{2}$. Widening the gap between the two decisions promotes a "faster" adoption of the optimal behavior. Thus, only $\alpha > \frac{1}{2}$ accelerates the evolutionary process.[7]

Lemma 2 *Given assumptions (1) and (2), the Concorde fallacy accelerates the evolutionary process for a given initial condition if and only if $\alpha > \frac{1}{2}$:*

$$p(t, C) < p(t, D) \Leftrightarrow \alpha > \frac{1}{2}.$$

Proof: See Appendix.

This highlights a key feature of the underlying structure. As argued above, any sunk cost problem must involve at least two interrelated decision problems. Committing the Concorde fallacy means doing wrong thing with respect to the second problem, or generally, in some situations. The gain may consist in learning to get things right in other, more important situations. The Concorde fallacy, however, does not always teach the right lesson. In our game, in "good"

[7] Note that "faster" does not mean that dp/dt increases by lowering C. This cannot be true for all values of t since p converges to 1 for all values of C. Thus, if the initial "speed" dt/dC is larger for $\widehat{C} < C$ in the beginning, it must be slower at some other point in time.

states of the world it does: it reduces the payoff whenever the wrong first period decision had been taken, and thus subjects adapt faster to what to do in the good state of the world. In the bad state of the world, the sunk cost fallacy decreases the payoff of the correct first period choice I_A, therefore hindering the process towards the optimal decision. With $\alpha < \frac{1}{2}$, the latter occurs more often, thus, $\alpha > \frac{1}{2}$ is a necessary condition for a positive evolutionary effect of committing the sunk cost fallacy. Throughout the paper, $\alpha > \frac{1}{2}$ is assumed.

5 Results for different adaptation environments

5.1 Biological evolution and the sunk cost problem

From the perspective of evolutionary biology, a typical question to ask in this context is: Can committing the Concorde fallacy increase the growth rate π, defined in (5), of a population in time T, in particular for large values of T? The behavior with the highest growth rate will drive out any other behavior in the long run. Figure 3 illustrates that in the long run the growth rate π will be larger when a population does not commit the Concorde fallacy. In this example $A = 8 > C = 5 > B = 2 > D = 1.5$; $\alpha = 0.96$, and the initial condition is $A_0 = 1$, i.e. at $t = 0$ half of the population chooses I_A, and the other half chooses I_B. The thin line reflects the population growth rates of a behavior that avoids the Concorde fallacy. The bold line reflects the population growth rates of a population committing the Concorde fallacy, namely, earning $D = 1.5$ instead of $C = 5$. The population that commits the Concorde fallacy starts with a lower growth rate at $t = 0$. It then "overtakes" for some intermediate period, i.e. there is a time interval in which the population committing the Concorde fallacy grows faster. In the long run, the growth rates of both populations converge, but the growth rate of the Concorde fallacy population converges to a lower level.

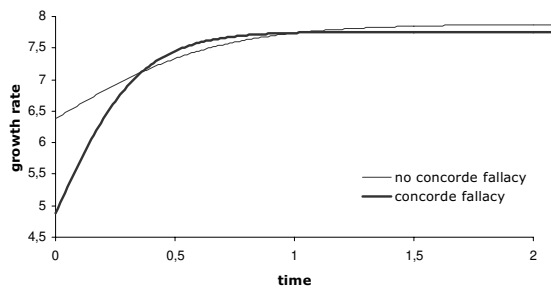

Figure 3: Biological Evolution and the Concorde Fallacy

The properties of this example can be generalized as stated in the following proposition.

Proposition 1 *A small deviation from the correct second stage decision, which yields some* $D = C - \varepsilon$, *will have the following effect on the growth rate* $\pi(t)$: *(i) The growth rate will be lower in the beginning, i.e. for* t *close to zero* $\pi(t, C) > \pi(t, D)$. *(ii) The growth rate will be lower in the long run: For* $t \to \infty$ *again* $\pi(t, C) > \pi(t, D)$. *(iii) For intermediate values of* t, *the growth rate will be higher for* α *sufficiently close to* 1 *and* $(A - B)$ *sufficiently large.*

Proof: See Appendix.

For extreme values of t, the intuition for the result is straightforward. In either extreme, the population states p are the same for both kinds of behavior. They start at the same point, and for large values of t both will be close to 1. If the probability of avoiding the wrong first decision is the same, committing the sunk cost fallacy must yield a lower payoff due to the second stage mistake. In-between, committing the Concorde fallacy might result in a sufficient advantage in the speed of learning so that committing the Concorde fallacy can provide a lead for some time interval. This can happen only if the good state of the world is highly likely, i.e. α is close to 1. As mentioned already, only in the good state of the world does the punishment effect of the Concorde fallacy drive evolution in the right direction. What the Concorde fallacy delivers is a faster transition to a high p. The prize to be gained from that is exactly the difference between A and B, the difference between the payoffs of selecting the right initial option I_A and the wrong initial option I_B.

Proposition 1 provides a clear answer to the biologist's question: In the long run, avoiding the Concorde fallacy yields higher growth rates. Thus, for evolutionary processes where the relevant time horizon is sufficiently large, evolution will drive out the Concorde fallacy.

What about the higher intermediate growth rates? Not all biological processes might have enough time to reach the long-run state in which committing the Concorde fallacy again yields lower growth rates. Proposition 2 states that, although the *growth rate* might be higher at some points in time, the *population size* of the Concorde fallacy population will never be higher.

Proposition 2 *The population size of a population committing the Concorde fallacy by earning* $D < C$ *will never exceed the population size of a population not committing the Concorde fallacy.*

Proof: See Appendix.

The higher evolutionary speed is not able to compensate for the lower reproduction rates in the beginning. Thus, in this framework, committing the Concorde fallacy is driven out by the biological evolutionary process, a result in conformity with the biological evidence that animals do not commit the Concorde fallacy.

5.2 Individual (reinforcement) learning

A change of perspective is required when turning to the individual reinforcement learning process. For human decision making, it is realistic to assume that we do

not optimize in every situation but use "Simple heuristics that make us smart" (Gigerenzer, Todd, and the ABC Research Group (1999)). To stick to the initial decision is a simple heuristic; but does the Concorde fallacy make us "smart", i.e. is it in some sense beneficial? In our context, this can be rephrased in the following form: Will the utility of a subject over its finite lifetime T increase when it ex-ante commits to the Concorde fallacy in order to speed up learning? Formally, this means that we want to compare the expected payoffs at $t = 0$ for the time period until T for the cases in which the Concorde fallacy is committed to those in which it is not.

Since payoffs in this context are not meant as offspring but as some sort of consumption (or monetary payoffs that allow for consumption), it is interesting to allow for risk aversion. We will therefore use a utility function $u(\cdot)$, with $u' > 0$ and $u'' < 0$. Since risk aversion will play an important role, comparative statics with respect to risk aversion are of particular interest. We therefore specify the utility function to exhibit constant absolute risk aversion a:[8]

$$u(x) = -e^{-ax}, \ a > 0. \tag{11}$$

Again, we first derive the result for not committing the Concorde fallacy and then ask whether switching to a behavior which yields $D = C - \varepsilon$ instead of C can increase the expected utility. The expected utility from not committing the concord fallacy is given by:

$$EU(\pi(t)) = \int_0^T \left(\frac{e^{zt}}{e^{zt} + A_0} x + \frac{A_0}{e^{zt} + A_0} y \right) dt, \tag{12}$$

where:

$$\begin{aligned} z &= \frac{\alpha(A - C) + (1 - \alpha)(C - B)}{A}, \\ x &= -\alpha e^{-aA} - (1 - \alpha) e^{-aC}, \\ y &= -\alpha e^{-aC} - (1 - \alpha) e^{-aB}. \end{aligned}$$

[8] Note that utility is defined as a negative value, thus we could interpret the problem as one of minimizing the disutility of the decision, or we simply re-scale the utility function by adding a sufficiently large arbitrary constant to get positive values for the utility. None of the results would be affected by such a re-scaling, since we focus on the difference of utilities only.

This can be rewritten (see the proof of Proposition 2 for details) to get:[9]

$$EU\left(\pi\left(p\right)\right)=\frac{x\ln\left(e^{zT}+A_0\right)-y\ln\left(A_0e^{-zT}+1\right)+\left(y-x\right)\ln\left(1+A_0\right)}{z}. \quad (14)$$

We are interested in the question whether the expected utility can be increased when reducing the payoff C to some $D=C-\varepsilon$. Figure 4 shows a case where this actually happens for intermediate values of T. The parameters of the example are: $A=1.2$, $C=1$, $B=-1.8$, $D=0.7$, thus, $\varepsilon=0.3$, i.e. in the case where switching in the second stage is optimal, committing the Concorde fallacy yields a lower payoff of 0.7 instead of 1. The probability of the good state of the world is $\alpha=0.96$, the individual starts with a 50-50 probability of selecting I_A or I_B, and it is risk averse with $a=1$. The properties of the functions are the same as in the biological application. Committing the Concorde fallacy yields lower payoffs in the beginning and in the end (thin line), while it might yield higher payoffs for an intermediate time period.

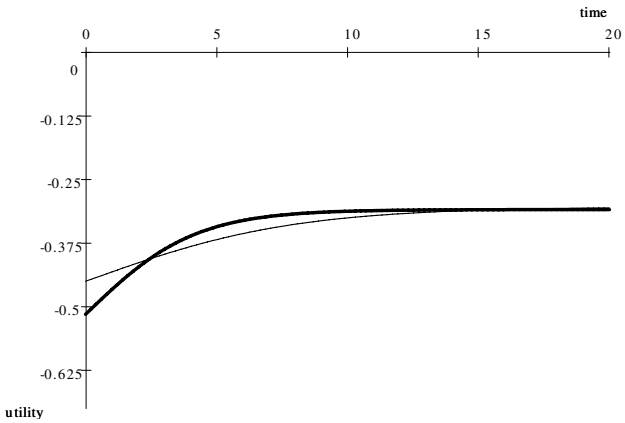

Figure 4: Expected utilities over time. Thin = avoiding Concorde fallacy; bold = committing concorde fallacy

However, analyzing the payoff difference yields a surprising result. Figure 5 depicts the difference between the two payoff functions of Figure 4 (the utility

[9] (14) is a convex function of the time horizon T :

$$\begin{aligned}\frac{\partial EU}{\partial T} &= \frac{e^{zT}}{e^{zT}+A_0}x+\frac{A_0}{e^{zT}+A_0}y<0, \text{since } x,y<0 \text{ and}\\ \frac{\partial^2 EU}{\partial T^2} &= zA_0\frac{e^{zT}\left(x-y\right)}{\left(e^{zT}+A_0\right)^2}>0, \text{ since } x>y. \end{aligned} \quad (13)$$

The expected utility goes to $-\infty$ for $T\to\infty$, but with a decreasing rate:

$$\begin{aligned}\lim_{T\to\infty}\frac{\partial EU}{\partial T} &= x,\\ \lim_{T\to\infty}\frac{\partial^2 EU}{\partial T^2} &= 0.\end{aligned}$$

of not committing the Concorde fallacy minus the utility of committing the Concorde fallacy). Since the integral of this function equals the expected utility difference over time, it is obvious that for some T, committing the Concorde fallacy is worthwhile.

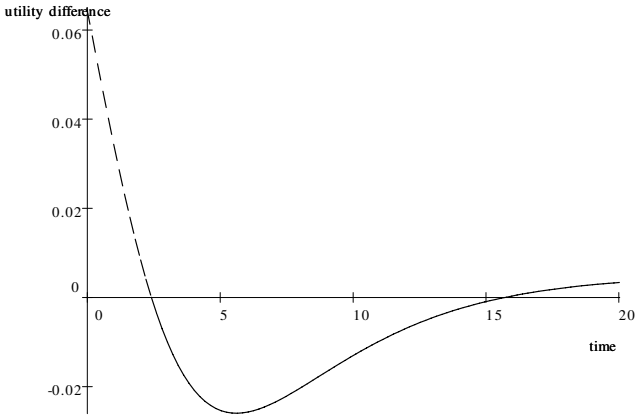

Figure 5: Difference in Expected Utilities

If the subject lives for $T = 20$, it can expect a utility of $EU^{NC} = -6.867$ when avoiding the Concorde fallacy, while it can expect $EU^{C} = -6.768$ if it commits the Concorde fallacy. Thus, the fact that occasionally committing the Concorde fallacy speeds up the learning process, eventually leading to fewer mistakes in the first decision, overcompensates for the lower utility whenever the second decision it taken incorrectly by 0.1. The following proposition states sufficient conditions for the occurrence of this beneficial effect from committing the Concorde fallacy.

Proposition 3 *For any time interval T a decision-maker with constant absolute risk aversion a can increase its expected utility by committing the Concorde fallacy, i.e. earning $D = C - \varepsilon$ instead of C, if B is sufficiently small and a and α are sufficiently large.*

Proof: See Appendix.

Proposition 3 stands in strong contrast to Proposition 2. While for a biological application, the Concorde fallacy can never be regarded as superior, with humans we can always find conditions such that this is the case. The key for explaining the difference in the results lies in the assumption of risk aversion. Committing the Concorde fallacy increases the evolutionary speed in both contexts. More quickly learning to take the optimal initial decision also produces a more steady stream of payoffs, since it more rapidly leads to the avoidance of the bad payoff B. The cost of this is the lower expected payoff in the long run. Risk aversion, however, attributes an extra benefit to lower variance of payoffs which might compensate for the loss from a lower average return.

The conditions in Proposition 3 highlight the importance of the risk aversion assumption for the results. The more risk averse a decision-maker is, the more he or she wants to avoid a very bad outcome even if it is very rare. This is what the conditions in the proposition say: B needs to be very small, that is, the "worst case" is indeed very bad; α needs to be close to unity, which means that this very bad outcome is very rare; and a sufficiently large means that the decision-maker is indeed very risk averse.

5.3 Social (imitation) learning

The formalism for the imitation rule is identical to the formalism of the reinforcement rule. If we again assume the specification of risk averse players, all formal results carry over to this context. This means that in a population in which everybody uses a imitation learning rule, the expected utility of each member of the population can be increased by committing the Concorde fallacy for the parameter conditions stated in Proposition 3.

If everybody abstains from adapting to the state of the world, the learning process of the population will be accelerated. A subject will be more likely to choose the right decision in the next period due to the "voluntary" payoff reduction (or self-punishment): She will observe a lower payoff from the wrong decision in the good state of the world (D instead of C), which will drive the subject to avoid the wrong decision more rapidly.

This situation is, however, not stable against an invasion of a small number of players who follow the same learning rule, but for themselves always choose the correct second stage decision. As long as the number of such deviators is small, the learning speed will not be reduced significantly. There will be enough other players from whom to learn the right behavior, but since one always does better than these players, such a behavior yields superior profits. This will undermine the learning dynamics, since these deviators also earn relatively high profits in the case where they made the wrong initial decision. Even more, if the deviation itself can be imitated, in the long run everybody will deviate towards not committing the Concorde fallacy, since this always yields higher profits. The population thus looses the benefit from faster learning.

This highlights that, in an imitation learning process, committing the Concorde fallacy produces a positive externality. This is similar to the externality produced, for example, by not crossing the street when the traffic lights are red although there is no traffic in order not to set a bad example for children. Thus, for the population to protect the social (learning) gains derived from committing the Concorde fallacy, there must be some punishment for deviation. In the traffic light example: from time to time one hears unfriendly comments when crossing the street when the light is red – which is some sort of social punishment. Being regarded as somebody who "wastes" resources when avoiding the Concorde fallacy might have a similar effect.

6 Discussion

This paper has tried to provide an evolutionary explanation for why humans care about sunk cost while animals don't. Sunk cost can serve as an instrument for self-punishment in case of wrong initial decisions. This accelerates learning and evolutionary selection towards optimal initial decisions at the cost of suboptimal follow-up decisions. Committing the Concorde fallacy unambiguously reduces payoffs in the long run. Thus, populations that avoid the Concorde fallacy in the long run will dominate those which do not. Even more, the expected population size will be larger at any point in time when the Concorde fallacy is avoided. Humans, however, might care less about expected payoffs than expected utility. Committing the Concorde fallacy not only reduces expected payoffs, but also the variance of payoffs. Thus, a risk averse individual might benefit from committing the Concorde fallacy.

In order to capture the idea of a sunk cost behavior, in the payoff matrix of Table 1, the payoffs for the combinations of initial decisions and the states of the world (I_A, θ_B) and (I_B, θ_A) were identical (namely C). We analyzed only a simultaneous variation of the payoff for both outcomes. The formal analysis nevertheless highlights that a downward manipulation of payoffs can be beneficial in general in adaptation setups. If one were allowed to reduce any payoff from the four different combinations, a risk averse subject would probably like to reduce only the payoff of (I_B, θ_A). Choosing the optimal level of this payoff would then correspond to finding the optimum (self-)punishment.

We have just briefly discussed the stability issues for the imitation learning process. Further research would be needed to clarify theoretically as well as empirically under which circumstances humans are willing to support a Concorde fallacy regime by punishing those who do not commit the Concorde fallacy.

We have simplified our discussion by assuming risk aversion for humans and risk neutrality for biological selection processes. The former seems less problematic than the latter. There is an extensive discussion in the literature of theoretical as well as experimental evolutionary biology on the issue of risk aversion. It is well-established that if the extinction of a gene (or more precisely: an allele) is a real threat, evolution will select risk averse behavior. Philippi and Seger (1989) provide an introduction to the biological literature on "bet-hedging" (while further theoretical issues have been discussed in Grafen (1999)). This literature suggests that animals in populations with a large number of individuals with the similar genes, e.g. social insects like ants or bees, are almost risk neutral in their individual behavior (since extinction of a gene is a very unlikely event), while others, where individuals differ more with respect to their genetic endowment, behave more risk averse. This could, however, offer the chance to test whether risk aversion plays a role. At first sight, this thinking harmonizes with our hypothesis: Humans differ more with respect to their genes than insects, birds, or fish do. And it is among these sorts of "lower animals" that the absence of the Concorde fallacy has been found.

7 Appendix

7.1 Proof of Lemma 2

Compare two different situations with the same initial condition p_0: In situation 1, in the underlying decision problem, subjects receive C; in situation 2 subjects receive $D < C$. Both situations define evolutionary processes, $p(t, C)$ and $p(t, D)$, which cannot intersect: At $p(0)$, $p(t, D)$ will have the larger growth rate if and only if $\alpha > \frac{1}{2}$. Now assume that at some $\widehat{t} > 0$ the two functions intersect, which would imply that $p\left(\widehat{t}, C\right) = p\left(\widehat{t}, D\right)$ and $\frac{\partial p\left(\widehat{t},C\right)}{\partial t} > \frac{\partial p\left(\widehat{t},D\right)}{\partial t}$, which cannot be the case for $\alpha > \frac{1}{2}$. Thus, $p(t, C) \leq p(t, D)$ $\forall t > 0$ for $\alpha > \frac{1}{2}$. ■

7.2 Proof of Proposition 1

The starting point is the growth rate π that follows from the optimum second stage choice, yielding C whenever "switching" in the second stage (choosing i_B after I_A, or i_A after I_B) is optimal. We compare this to a "non-switching" behavior by analyzing the partial derivative with respect to C.

$$\begin{aligned}\frac{\partial \pi}{\partial C} &= \frac{A_0 z' t e^{zt} x + x' e^{2zt} + A_0 x' e^{zt} + A_0 y' e^{zt} + A_0^2 y' - A_0 y z' t e^{zt}}{\left(e^{zt} + A_0\right)^2} \\ &= \underbrace{\frac{A_0 z' t e^{zt} (x - y)}{\left(e^{zt} + A_0\right)^2}}_{<0} + \underbrace{\frac{x' e^{zt} + y' A_0}{e^{zt} + A_0}}_{>0}. \end{aligned} \tag{15}$$

For the Concorde fallacy to increase π, this expression must be negative, requiring:

$$\begin{aligned} A_0 (2\alpha - 1) t e^{zt} \alpha (A - B) &> \left((1 - \alpha) e^{zt} + \alpha A_0\right)\left(e^{zt} + A_0\right), \\ A_0 \left[(2\alpha - 1) t e^{zt} \alpha (A - B) - \alpha A_0 e^{zt}\right] &> e^{2zt} (1 - \alpha). \end{aligned} \tag{16}$$

For $t \to 0$, the left hand side of (16) goes to $-\left(\alpha A_0^2 + 1\right) < 0$, while the right hand side goes to $(1 - \alpha) > 0$, a contradiction, which proves (i).

(16) can be rewritten to yield:

$$A_0 \left[1 - \frac{\alpha A_0}{e^{zt} t (2\alpha - 1) \alpha (A - B)} - \frac{1}{(2\alpha - 1) t \alpha (A - B)}\right] > \frac{e^{zt}}{t} \frac{(1 - \alpha)}{(2\alpha - 1) \alpha (A - B)}. \tag{17}$$

For $t \to \infty$, the left hand side of (17) goes to A_0, some positive constant, while the right hand side is positive and goes to infinity ($\lim_{t\to\infty} e^{zt}/t = \lim_{t\to\infty} z e^{zt}/1$ by l'Hospital's rule), which again yields a contradiction, proving (ii). Finally, for $\alpha \to 1$, and by substituting for $\lim_{\alpha\to 1} z = (A - C)$, (17) becomes:

$$\begin{aligned} A_0 \left[1 - \frac{A_0}{e^{(A-C)t} t (A - B)} - \frac{1}{t (A - B)}\right] &> 0, \\ e^{t(A-C)} \left[t (A - B) - 1\right] &> 1, \end{aligned}$$

which will be fulfilled for $A - B$ sufficiently large, proving (iii). ■

7.3 Proof of Proposition 2

First note that:

$$\int_0^T \pi(t)\,dt = \int_0^T \frac{dN_t}{dt}\frac{1}{N_t}dt = [\ln N_t]_0^T = \ln N_T - \ln N_0.$$

Thus, the larger the integral from 0 to T, the larger the population size N_T in T, and therefore, if decreasing C to $D = C - \varepsilon$ increases the integral, it also increases the population size:

$$\int_0^T \pi(C,t)\,dt < \int_0^T \pi(D,t)\,dt \Leftrightarrow N_T(C) < N_T(D).$$

Call

$$EU(T,C) = \int_0^T \pi(t,C)\,dt.$$

To prove the proposition, it is sufficient to show that

$$\frac{\partial EU}{\partial C} > 0.$$

The proof is by contradiction. Using (5) for π and the solution to the ordinary difference equation for $p(t)$ in (9) we can write:

$$\begin{aligned} EU(\pi(t)) &= \int_0^T \left(\frac{e^{zt}}{e^{zt}+A_0}x + \frac{A_0}{e^{zt}+A_0}y\right)dt \\ &= \int_0^T \left(\frac{e^{zt}}{e^{zt}+A_0}x + \frac{A_0e^{-zt}}{1+A_0e^{-zt}}y\right)dt. \end{aligned}$$

In order to use that for $\gamma = \ln f(x)$ we have $\gamma' = \frac{f'(x)}{f(x)}$, this expression needs to be expanded by z and $-z$, respectively, to get:

$$EU = \frac{x}{z}\int_0^T \frac{ze^{zt}}{e^{zt}+A_0}dt + \frac{y}{-z}\int_0^T \frac{-zA_0e^{-zt}}{1+A_0e^{-zt}}dt, \tag{18}$$

since

$$\begin{aligned} \frac{\partial}{\partial t}\left(e^{zt}+A_0\right) &= ze^{zt} \text{ and} \\ \frac{\partial}{\partial t}\left(1+A_0e^{-zt}\right) &= -zA_0e^{-zt}, \end{aligned}$$

(18) becomes:

$$\begin{aligned}
EU &= \frac{x}{z}\left[\ln\left(e^{zt}+A_0\right)\right]_0^T - \frac{y}{z}\left[\ln\left(1+A_0e^{-zt}\right)\right]_0^T \\
&= \frac{x}{z}\left[\ln\left(e^{zT}+A_0\right)-\ln\left(1+A_0\right)\right] \\
&\quad -\frac{y}{z}\left[\ln\left(1+A_0e^{-zT}\right)-\ln\left(1+A_0\right)\right] \\
&= \frac{x\ln\left(e^{zT}+A_0\right)-y\ln\left(A_0e^{-zT}+1\right)+(y-x)\ln\left(1+A_0\right)}{z}. \qquad (19)
\end{aligned}$$

Taking the derivative of (19) with respect to C yields (where x', y', z' denote partial derivatives with respect to C):

$$\begin{aligned}
\frac{\partial EU}{\partial C} &= -\frac{z'}{z^2}\left[x\ln\left(e^{zT}+A_0\right)-y\ln\left(A_0e^{-zT}+1\right)+(y-x)\ln\left(1+A_0\right)\right] \\
&\quad +\frac{1}{z}\left[x'\ln\left(e^{zT}+A_0\right)+x\frac{z'Te^{zT}}{e^{zT}+A_0}-y'\ln\left(A_0+1\right)+y\frac{z'TA_0e^{-zT}}{A_0e^{-zT}+1}+\ln\left(1+A_0\right)\left(y'-x'\right)\right] \\
&= \frac{1}{z^2}\left[\ln\left(e^{zT}+A_0\right)\left(x'z-z'x\right)+\ln\left(1+A_0e^{-zT}\right)\left(y'z-z'y\right)\right. \\
&\quad +\ln\left(1+A_0\right)\left(y'z-yz'+xz'-x'z\right) \qquad (\spadesuit) \\
&\quad \left.+z'zT\frac{e^{zT}x+A_0y}{A_0+e^{zT}}\right].
\end{aligned}$$

Straightforward calculations – using the definitions for x, y, and z – show that the term in line ($\spadesuit$) is zero. $\frac{\partial EU}{\partial C} < 0$ therefore requires:

$$\ln\left(e^{zT}+A_0\right)\left(x'z-z'x\right)+\ln\left(1+A_0e^{-zT}\right)\left(y'z-z'y\right) < -z'zT\frac{e^{zT}x+A_0y}{A_0+e^{zT}}. \qquad (20)$$

Since $z' = 1-2\alpha < 0$, both sides of the inequality are positive. And since $\ln\left(e^{zT}+A_0\right) \geq zT$, we obtain a lower bound for the left hand side in (20):

$$zT\left(x'z-z'x\right) < \ln\left(e^{zT}+A_0\right)\left(x'z-z'x\right)+\ln\left(1+A_0e^{-zT}\right)\left(y'z-z'y\right).$$

Given this lower bound, a necessary condition for (20) to hold is:

$$\begin{aligned}
zT\left(x'z-z'x\right) &< -z'zT\frac{e^{zT}x+A_0y}{A_0+e^{zT}} \\
z'x\left(\frac{e^{zT}}{e^{zT}+A_0}-1\right)+x'z+ &< -z'y\frac{A_0}{e^{zT}+A_0} \\
z'\frac{A_0}{e^{zT}+A_0}\left(y-x\right)+x'z &< 0,
\end{aligned}$$

a contradiction, since $z' < 0$ and $y < x$. ■

7.4 Proof of Proposition 3

Differentiating (12) with respect to C yields:

$$\begin{aligned}\frac{\partial EU}{\partial C} &= \int_0^T \frac{z'te^{zt}\left(e^{zt}+A_0\right)-z'te^{zt}e^{zt}}{\left(e^{zt}+A_0\right)^2}+x'\frac{e^{zt}}{e^{zt}+A_0}-\frac{z'te^{zt}}{\left(e^{zt}+A_0\right)^2}y+\frac{A_0}{e^{zt}+A_0}y'dt \\ &= \int_0^T z'\frac{e^{zt}}{\left(e^{zt}+A_0\right)^2}t\left(A_0x-y\right)+\frac{e^{zt}x'+A_0y'}{e^{zt}+A_0}dt. \qquad (21)\end{aligned}$$

We want to show that (21) is negative for B sufficiently low. Note that $B \to -\infty$ implies $z \to \infty$. Therefore, the second term in the integral goes to a positive constant x'. The limit of the first term in the integral is not straightforward:

$$\begin{aligned} & tz' \lim_{B\to-\infty}\frac{e^{zt}}{\left(e^{zt}+A_0\right)^2}\left(A_0x-y\right), \\ = \; & tz' \lim_{B\to-\infty}\frac{e^{zt}A_0x}{\left(e^{zt}+A_0\right)^2}-tz' \lim_{B\to-\infty}\frac{e^{zt}y}{\left(e^{zt}+A_0\right)^2}.\end{aligned}$$

While the first term goes to zero, in the second the nominator and denominator go to infinity for $B \to -\infty$. Applying l'Hospital's rule twice for the second term yields:

$$\begin{aligned} -tz' \lim_{B\to-\infty}\frac{e^{zt}y}{\left(e^{zt}+A_0\right)^2} &= -tz' \lim_{B\to-\infty}\frac{e^{zt}\left(-\alpha e^{-aC}-(1-\alpha)e^{-aB}\right)}{\left(e^{zt}+A_0\right)^2} \\ &= -tz' \lim_{B\to-\infty}\frac{t\alpha e^{-aC}+\left[(1-\alpha)t+a\right]e^{-aB}}{2te^{zt}+2tA_0} \\ &= -tz' \lim_{B\to-\infty}\frac{-a\left[(1-\alpha)t+a\right]}{2t^2(1-\alpha)}e^{-aB-zt},\end{aligned}$$

which will go to -∞ for $B \to -\infty$ if:

$$\begin{aligned} -aB-zt &> 1 \\ -aB-t\left(\alpha\left(A-C\right)+(1-\alpha)\left(C-B\right)\right) &> 1 \\ -aB+t(1-\alpha)B &> t\left(\alpha\left(A-C\right)+(1-\alpha)C\right) \\ B\left(t(1-\alpha)-a\right) &> t\left(\alpha\left(A-C\right)+(1-\alpha)C\right).\end{aligned}$$

This condition will be satisfied for $B \to -\infty$ if:

$$t(1-\alpha) < a,$$

which holds for any t if α is close to 1 and/or a is sufficiently large, as stated in the proposition. ■

References

Arkes, H. R., and P. Ayton (1999): "The Sunk Cost and Concorde Effects: Are Humans Less Rational Than Lower Animals?," *Psychological Bulletin*, 125(5), 591–600.

Arkes, H. R., and C. Blumer (1985): "The Psychology of Sunk Cost," *Organizational behavior and human decision processes*, 35, 124–140.

Boot, A. W. A. (1992): "Why Hang on to Loosers? Divestiture and Takeovers," *Journal of Finance*, 47(4), 1401–1423.

Curio, E. (1987): "Animal Decision-Making and the Concorde Fallacy," *Trends in ecology and evolution*, 2(6), 148–152.

Dawkins, R., and H. J. Brockman (1980): "Do Digger Wasps Commit the Concorde Fallacy?," *Animal Behavior*, 28, 892–896.

Edwards, C. (1973): "A Cost-Benefit Analysis of the Concorde Project: A Comment," *Journal of Transport Economics and Policy*, 7(3), 300–301.

Garland, H., and S. Newport (1991): "Effects of Absolute and Relative Sunk Costs on the Decision to Persist with a Course of Action," *Organizational Behavior and Human Decision Processes*, 48, 55–69.

Gigerenzer, G., P. M. Todd, and the ABC Research Group (1999): *Simple Heuristics That Make Us Smart*. Oxford University Press, New York.

Grafen, A. (1999): "Formal Darwinism, the Individual-as-Maximizing-Agent Analogy and Bet-Hedging," *Proceedings of the Royal Society, London*, 266, 799–803.

Kahneman, D., and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47(2), 263–291.

Kanodia, C., R. Bushman, and J. Dichhaut (1989): "Escalation Errors and the Sunk Cost Effect: An Explanation Based on Repuation and Information Asymmetries," *Journal of Accounting Research*, 27(1), 59–77.

McCarthy, A. M., F. D. Schoorman, and A. C. Cooper (1993): "Reinvestment Decisions by Entrepreneurs: Rational Decision Making or Escalation of Commitment?," *Journal of Business Venturing*, 8(1), 9–24.

Philippi, T., and J. Seger (1989): "Hedging One's Evolutionary Bets, Revisited," *Trends in ecology and evolution*, 4(2), 41–44.

Samuelson, L. (1997): *Evolutionary Games and Equilibrium Selection*. MIT Press, Cambridge, MA.

Schlag, K. (1998): "Why Imitate , and If so, How?," *Journal of Economic Theory*, 78(1), 130–156.

Sliwka, D. (2005): "Why Good Managers May Stick to Bad Decisions - On Managerial Turnover and Strategic Change," Mimeo, University of Cologne.

Thomas, R. (1981): *Microeconomic Applications: Understanding the American Economy*. Wadsworth, Belmont, CA.

Weibull, J. W. (1995): *Evolutionary Game Theory*. MIT Press, Cambridge, MA.

Woolley, P. K. (1972): "A Cost-Benefit Analysis of the Concorde Project," *Journal of Transport Economics and Policy*, 6(3), 224–239.

Zeelenberg, M., and E. van Dijk (1997): "A Reverse Sunk Cost Effect in Risky Decision Making: Sometimes We Have Too Much Invested to Gamble," *Journal of Economic Psychology*, 18, 677–691.