CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Discussion Paper No. 2014-06

Fabrizio Adriani and Silvia Sonderegger
May 2014

Evolution of similarity judgements in intertemporal choice

CeDEx Discussion Paper Series
ISSN 1749 - 3293

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/cedex for more information about the Centre or contact

Suzanne Robey
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0)115 95 14763
Fax: +44 (0) 115 95 14159
suzanne.robey@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx

Evolution of similarity judgements in intertemporal choice

By Fabrizio Adriani and Silvia Sonderegger*

We study Nature's trade-off when endowing people with the cognitive ability to distinguish between different time periods or different prizes. Our key premise is that cognitive ability is a scarce resource, to be deployed only where and when it really matters. We show that this simple insight can explain a number of observed anomalies: (i) time preference reversal, (ii) magnitude effects, (iii) cycles, (iv) interval length effects. An implication of our analysis is that, from an evolutionary perspective, people may be suffering from too much tendency to postpone (rather than to anticipate) consumption, turning upside-down existing interpretations of preference reversal.
JEL: D01, D81, D83

"*After all, tomorrow is another day.*"
Scarlett O'Hara.

Consider an individual who is confronted with two time horizons, t periods from now, and $t+1$ periods from now. How will he feel about them? Casual introspection suggests that, if t is sufficiently large, he will see the two horizons as very similar, to the point of being indistinguishable. For instance, it is unlikely that he will perceive any difference between an horizon of 1 year and

* Adriani: Department of Economics, University of Leicester, fa148@le.ac.uk. Sonderegger: Department of Economics, University of Nottingham and CeDEx, silvia.sonderegger@nottingham.ac.uk. We thank Alberto Bisin, Subir Bose, Robin Cubitt, Sergio Currarini, Eddie Dekel, Péter Esö, Luis Rayo, Larry Samuelson, Daniel Seidmann, Balázs Szentes, Chris Wallace, Jörgen Weibull, Eyal Winter and Piercarlo Zanchettin for useful comments and discussions. All errors are obviously our own.

one of 1 year and 1 day from now. By contrast, he will presumably see a difference between today and tomorrow.

Intertemporal choice and intertemporal preferences have been the object of much research.[1] The experimental literature has identified a number of "anomalies," some of which can be considered stylized facts. These include:

1) Time preference reversal: The rate of time preference appears to decline with time,

2) Magnitude effects: Subjects appear to be more patient towards larger rewards,

3) Cycles: Intertemporal choices are often intransitive,

4) Interval length effects: The average discount rate for a period of time might differ from the rate resulting from compounding the average rates of different subperiods.

Do these phenomena have a common denominator, and, if yes, what is their underlying *ultimate* common cause? Is it related to the casual observation that individuals may fail to perceive differences in some situations but not in others? Why did *homo sapiens* fail to evolve a mechanism to overcome these inconsistencies? To our knoweldge, the literature has failed to provide a unified account that allows to answer these questions. In this paper, we try to fill this gap by developing an evolutionary theory of similarity judgements.

The notion that similarity judgements may matter for intertemporal choice dates back to Leland (2002) and Rubinstein (2003). These works, however, fall short of providing a theory of *how* similarity relations emerge and *why*

[1] Loewenstein and Prelec (1992), Frederick et al. (2002) and Manzini and Mariotti (2009) provide comprehensive surveys of the literature.

they should take a particular form. In this paper, we adress these questions, by going one step back. We consider the trade-off faced by Nature when endowing an individual with the ability to distinguish between different time periods or between different quantities. On the one hand, better ability to distinguish allows the individual to take better decisions. On the other hand, it also requires greater cognitive ability, which comes at a cost.

There is a large literature in biology and anthropology on the evolutionary trade-off between greater cognitive ability and the high energy demands of a larger brain, see e.g. Aiello and Wheeler (1995). Bigger brains also generate other types of evolutionary disadvantages – for instance, child birth becomes more hazardous.[2] These observations underpin our key premise, namely that cognitive ability is a scarce resource, to be deployed only where and when it really matters. The contribution of this paper is to show that this simple insight allows to shed light on the possible ultimate causes of anomalies (1)-(4). In this respect, our work complements the descriptive (or proximate) accounts of these phenomena that have been provided by the literature.[3]

Our characterization of the solution to Nature's problem proceeds in two steps. In the first step, we show that Nature will not find it optimal to endow individuals with the ability to distinguish between different time periods (or different prizes) if these are sufficiently similar. In other words, some pooling will occur. In the second step, we show that this pooling process exhibits a number of regularities, which can then be used to make empirical predictions.

To build intuition on how pooling operates, consider for instance the tools provided by languages to describe different "shades" of time. Italian distinguishes between remote and recent past, present and future, while Kalaw

[2]See e.g., Lovejoy (1988).

[3]See e.g. Binmore (2005) for a discussion of the relationship between ultimate evolutionary causes and proximate psychological mechanisms.

Lagaw Ya (an Australian Aboriginal language, which allegedly possesses the largest number of tenses) has remote past, recent past, today past, present, near future and remote future. Even this language, though, exhibits a degree of coarseness. Although it would be simplistic to assume that these categories fully describe the speaker's perception of time, this suggests that some bunching may naturally take place when considering time periods.

More formally, we show that (under some restrictions) the solution to Nature's problem takes the following form: For each time t there exists an interval around t such that all time periods within that interval are perceived as indistinguishable from t. Similarly, around each prize x there exists an interval such that all prizes belonging to that interval are indistinguishable from x. We call these intervals *similarity intervals* in the time and in the prize dimensions.

We find that similarity intervals exhibit a number of regularities. The first key property is that, in the time dimension, similarity intervals tend to *expand* as the time period t being considered is pushed further into the future, as depicted in Figure 1. This implies that an individual will be more likely to distinguish between today and tomorrow than between 1 year and 1 year and 1 day from now.

Intuitively, the gains that can be reaped from the ability to distinguish between distant time periods will only be realized in the distant future. Their *present* value in terms of enhanced fitness is therefore quite small. As a result, Nature does not find it worthwile to spend scarce resources to endow the individual with the ability to distinguish between these time periods. By contrast, this argument does not apply to time periods that are closer to the present.

We show that this feature of similarity intervals generates time preference reversal, in the form that has been documented empirically. When choosing between two bundles involving large consumption delays, the agent perceives

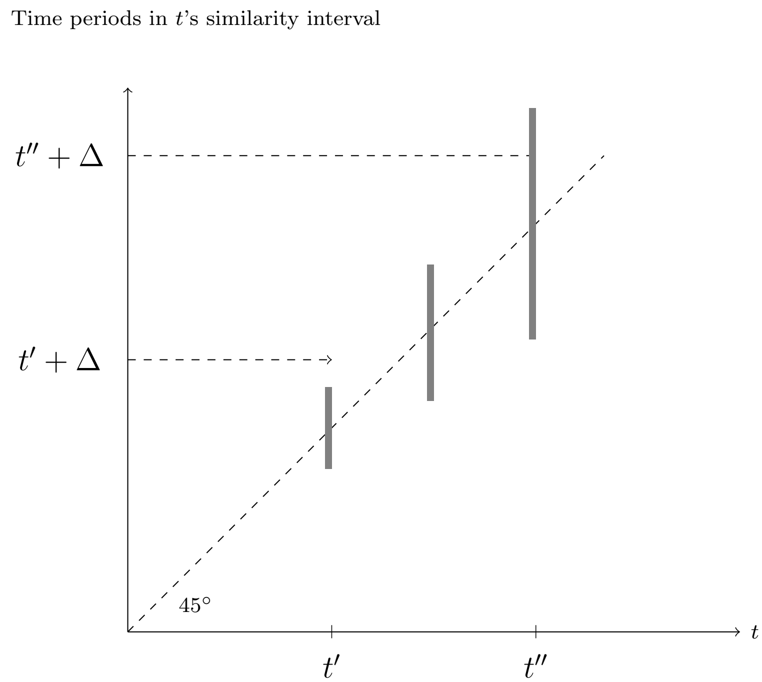

Figure 1. Similarity intervals expand as the time horizon increases. The individual perceives a difference between t' and $t' + \Delta$, but not between t'' and $t'' + \Delta$.

the two time periods as equally remote. His preferences over the two bundles are thus shaped by comparing prizes. By contrast, when faced with a choice between bundles involving small delays, the agent *can* see a difference between the time periods, and is therefore much more inclined to favor earlier consumption over a bigger prize. The important feature is that we are able to show that this is part of the solution to Nature's problem. Our story thus differs quite considerably from existing accounts.[4]

An interesting corollary of our analysis is that, for time preference reversal to

[4]Early formal accounts of the problem of time-inconsistency include Strotz (1956) and Phelps and Pollak (1968). Fudenberg and Levine (2006) argue that the behavior of individuals can be thought of as the outcome of the interactions of two subsystems (the "long-run self" and the "short-run self"). Leland (2002) and Rubinstein (2003) present accounts based on similarity relations; However, as we have argued, their models do not explain how similarity relations emerge or why they should take a specific form.

occur, the sooner/smaller bundle must be fitness maximizing. The individual then prefers early consumption when facing two short-run alternatives, but (inefficiently) favors late consumption when facing two long-run alternatives. This suggests that people may be suffering from *too much tendency to postpone* consumption, and has important implications. First, it contributes to the current debate on welfare behavioral economics, by questioning the conventional wisdom that sees most people suffering from a *present* bias. Much of the existing literature treats (implicitly or explicitly) the short-term preferences for immediate gratification – which contrast with the individual's long-term preferences – as an error. For instance, O'Donoghue and Rabin (1999) say:

> We feel the natural perspective in most situations is the "long-run perspective"–what you would wish now (if you were fully informed) about your profile of future behavior.[5]

We argue that, in evolutionary terms, the opposite may actually be true. This observation suggests a possible reason why evolution has not endowed us with the ability to commit. From an evolutionary viewpoint, the ability to reverse a (poor) early choice as the time of consumption approaches may be quite valuable.[6]

The second key property we identify concerns the prize dimension. We show that the solution to Nature's problem exhibits the following regularity: The agent is more likely to distinguish between a prize x and a prize rx (for some constant r) as x increases. In other words, similarity intervals in the prize

[5] This perspective is also taken in O'Donoghue and Rabin's (2003, 2006) work on sin taxes. See also Bernheim (2009) and Bernheim and Rangel (2009) for recent contributions on welfare behavioral economics, which stress the additional complexities that must be faced within that context.

[6] This does not imply that the ability to commit may not be valuable in other settings, such as for instance in the presence of strategic interactions.

dimension tend to *contract* as the prizes being considered increase, as depicted in Figure 2.[7]

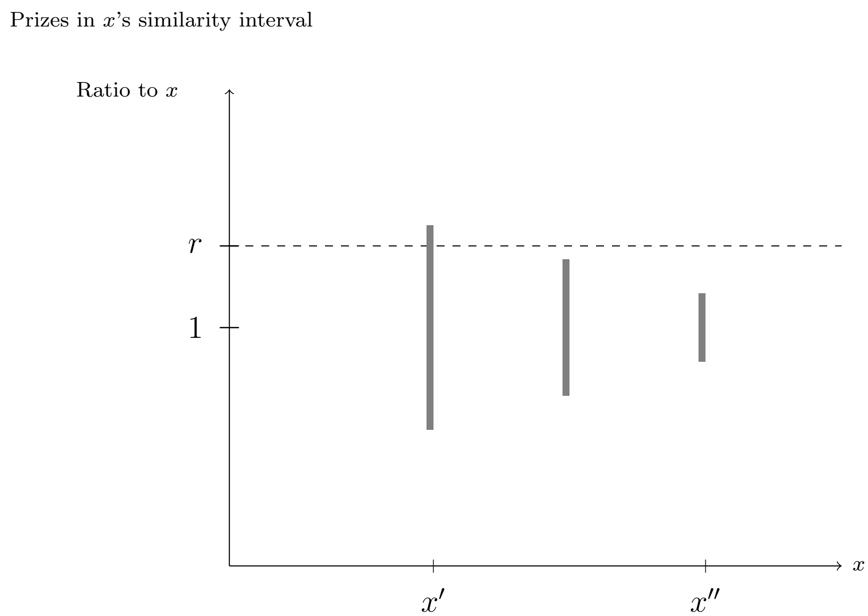

Figure 2. Similarity intervals contract as prizes become larger. The individual perceives a difference between x'' and rx'' but not between x' and rx'.

The rationale for this result is that larger prizes create larger gains to be reaped from making a better choice. Put differently, the loss that may be incurred by failing to make the distinction between x and rx gets larger for larger values of x – this is a direct consequence of the fact that the difference between x and rx increases in x. The implication is that, when x is large, Nature is more inclined to incur the necessary cost to distinguish between these two quantities. We show that this generates magnitude effects, in the form that has been observed empirically. Intuitively, when prizes (x and rx)

[7]Note that, in contrast with the case of time, here similarity intervals are defined in terms of the ratio to the prize being considered.

are small, the agent can't see much difference between them. His preferences over the two bundles are thus shaped by comparing consumption delays. By contrast, when prize magnitudes are large, the agent does distinguish between them, and is therefore much more inclined to delay consumption in order to obtain the larger prize.

Finally, we also show that our setup may generate cycles and interval length effects, in line with experimental evidence. In all the cases of preference reversals we characterize, we are able to single out which of the discordant preferences is "more efficient," in the sense of maximizing fitness.

The results described above are obtained in an environment with no in-built bias toward particular prizes or delivery times, i.e. we assume that Nature considers all delivery times and all prizes equally likely to occur. This helps to single out the driving forces at work. We next relax this assumption by considering more natural distributions for delivery times and prizes. In particular, we focus on the cases where delivery times follow a Poisson process and prizes have a unimodal distribution. Whenever short delays tend to be more frequent than long delays (as in a Poisson process) the qualitative features of the analysis remain unchanged –actually, the effect we describe is strengthened. A unimodal distribution of prizes would instead generate the following prediction: Similarity intervals would initially contract as the prize being considered increases (as in the case of a uniform distribution), but once the prize becomes sufficiently large they might eventually start to expand. The implication is that, while we should expect the magnitude effect to hold for small and intermediate quantities, it may fail when we consider very large prizes.

There is a growing interest within economics on the evolutionary foundations of preferences – see Robson and Samuelson (2011a) for a recent survey.[8] The

[8]Important contributions to the literature on the evolutionary foundations of preferences

most relevant contributions for our purposes are Robson (2001b), Rayo and Becker (2007) and Netzer (2009). Similar to us, these works model individual decision-making as the solution of an optimization problem in which Nature faces physiological constraints. However, differently from their approach, a key feature of our analysis is that individuals may be unable to tell the difference between similar time periods or similar prizes. A further difference with Robson (2001b) and Netzer (2009) is that, as our uniform example shows, our results do not rely on different objects occurring with different frequencies.

The paper is organized as follows. Section I presents the model, while Section II provides some general insights. Sections III and IV characterize the solution to Nature's problem when prizes and delivery times are drawn uniformly (and independently). Section V looks at the case of Poisson arrival times and unimodal prize distributions. In section VI we review the empirical predictions of existing descriptive models of time preferences, and discuss further related literature. Section VII concludes.

I. Setup

Environment We consider a decision maker who, at $t = 0$, must select between two (delivery time, prize) bundles (x_1, t_1) and (x_2, t_2). Prizes x_1 and x_2 are drawn from the same continuous distribution with full support $(0, \overline{x})$. We do not require them to be statistically independent, but assume that they are exchangeable and denote their joint density with $h(x_1, x_2)$.[9] Similarly, both delivery times are drawn from a continuous distribution with full support $(0, \bar{t})$ and are assumed to be exchangeable. We denote with $g(t_1, t_2)$ their joint den-

include Frank (1987), Waldman (1994), Robson (2001a), Samuelson (2004), Samuelson and Swinkels (2006), Robson and Samuelson (2009, 2011b), Alger and Weibull (2013).

[9]Two random variables are exchangeable if their joint probability remains the same when the two variables swap places, i.e., for all pairs (x_1, x_2), $h(x_1, x_2) = h(x_2, x_1)$.

sity. We will sometimes think of $\bar{t}$ and $\bar{x}$ as infinitely large, although there will be cases below where it will be convenient to assume finite upper bounds for prizes and delivery times.

Fitness We draw a conceptual distinction between individual fitness and individual preferences (which are addressed below). We assume that the fitness value of a given prize decays exponentially as the prize is moved further away into the future.[10] In particular, a prize x at time t is equivalent in terms of fitness to a prize $xe^{-\delta t}$ in the present, for some $\delta > 0$. As argued by Netzer (2009), there are many reasons for exponential fitness discounting, such as population growth (Hansson and Stuart 1999, Robson and Samuelson 2007) or declining fertility (Rogers 1994). The fitness-maximizing choice is thus (x_2, t_2) whenever $e^{-\delta t_1}x_1 < e^{-\delta t_2}x_2$, and is (x_1, t_1) whenever $e^{-\delta t_1}x_1 > e^{-\delta t_2}x_2$.

Similarity relations and Nature's problem The decision maker may or may not be able to distinguish between two time periods or between two prizes. Let us use the notation $a \approx b$ to indicate that the individual is unable to perceive the difference between two objects a and b and the notation $a \not\approx b$ to indicate that the individual perceives the difference.

Nature's task is to decide ex-ante (i.e. without knowing which bundles will occur) whether to endow the decision maker with the ability to distinguish between different time periods and different prizes. We assume that the ability to perceive a difference between any two objects (whether prizes or delivery times) involves a fixed fitness cost $c > 0$. We interpret c as the marginal cost of supporting a larger brain.[11] Hence, Nature's problem reduces to deciding,

[10]This is a natural benchmark. Moreover, it clarifies that our results hold *even when* fitness decays exponentially with time (rather than, say, decaying hyperbolically, in which case our results would appear less surprising). However, as should become clear below, any separable fitness function that is increasing in x and decreasing in t would deliver qualitatively similar results (although specific details may change).

[11]More precisely, c captures the present discounted value of the sum of evolutionary

for any (t_1, t_2) pair, whether it is worth paying this cost c in order to allow the individual to discriminate between t_1 and t_2, or not. The same applies to prizes. For any (x_1, x_2) pair, Nature decides whether it is worth paying this cost c in order to allow the individual to discriminate between x_1 and x_2, or not.[12]

Preferences We consider a setup where the only possible source of deviation from fitness maximization arises from the decision maker's inability to perfectly distinguish between times or prizes. This is clearly an abstraction but we believe it may be a natural starting point. We will accordingly assume that, so long as the decision maker is able to perceive the difference between two bundles along both dimensions (delivery times and prizes), his choice will reflect fitness maximization. However, if the individual perceives a difference along one dimension only, this dimension will determine his preferences, and this may generate a choice that differs from fitness maximization. In this respect, our description of the decision maker's preferences (and thus of his behavior) closely resembles the one given in Rubinstein (2003).

More precisely, consider a pair of bundles (x_1, t_1) and (x_2, t_2). We assume that

- if the decision maker is able to perceive a difference both between the de-

disadvantages imposed in each period by greater cognitive ability. Note that we implicitly assume that the cost of distinguishing between prizes and that of distinguishing between delivery times are the same. This is immaterial.

[12] An alternative approach would be to partition the time and quantity spaces into intervals such that the agent does not distinguish between objects located in the same interval. We have taken a different route here, for two reasons; (i) in the partition approach, the agent would be able to distinguish between two arbitrarily close objects (say, time periods t and $t + \Delta$, where Δ is small) if these lie on different sides of the bound of an interval, but would be unable to distinguish between two relatively distant objects if these lie on the same side. As will become clear below, this is suboptimal in our setting. (ii) Because of (i), the partition approach would not generate sharp predictions. In particular, it would be consistent with preference reversals occurring in either direction. By contrast, in our setup preference reversals may only take one specific form (which is the one consistent with the empirical evidence).

livery times and between the prizes of the two bundles, he prefers the bundle that generates the highest fitness, i.e. $\max\{x_1e^{-\delta t_1}, x_2e^{-\delta t_2}\}$, [If the two bundles generate identical fitness, we assume that he is indifferent and chooses either bundle with probability 1/2.]

• if he perceives a difference between delivery times but not a difference in prizes, he prefers the bundle that minimizes the waiting time, i.e. the bundle with delivery time $\min\{t_1, t_2\}$,

• if he perceives a difference between prizes but not a difference in delivery times, he prefers the bundle with the largest prize, i.e. $\max\{x_1, x_2\}$,

• if he perceives similarity along both dimensions, he is indifferent (and thus randomizes).

Table 1 summarizes the preferences of the individual according to what differences he is able to perceive. For the case $x_1 < x_2$ and $t_1 < t_2$, the fitness associated with his choice is also reported.

	$x_1 \not\approx x_2$	$x_1 \approx x_2$
$t_1 \not\approx t_2$	Bundle with max fitness $\max\left\{x_1e^{-\delta t_1}, x_2e^{-\delta t_2}\right\}$	Bundle with min waiting time $x_1e^{-\delta t_1}$
$t_1 \approx t_2$	Bundle with max prize $x_2e^{-\delta t_2}$	Indifferent $\frac{1}{2}\left(x_1e^{-\delta t_1} + x_2e^{-\delta t_2}\right)$

Table 1—Decision maker's preferences and associated fitness.

Notice that we are implicitly assuming that preferences are separable over the prize and time dimensions. There are a number of neuro-science papers that show that individuals evaluate these two aspects of a bundle separately. Using functional magnetic resonance imaging, Ballard and Knutson (2009) identify distinct patterns of brain activity associated with each dimension. Activation in the mesolimbic projection regions correlates with increasing the magnitude of future rewards, while activation in lateral cortical regions correlates with increasing delays of future rewards. Pine et al. (2009) present similar findings.[13]

II. Nature's trade-off

Nature faces a trade-off between endowing the individual with the ability to make finer distinctions between prizes and/or delivery times and the fitness cost of the capacity to make these distinctions. In this section, we analyze this trade-off.

When "designing" the brain of the individual, Nature cannot perfectly anticipate which pair of bundles the individual will be faced with. The ex-ante character of the problem implies that Nature will maximize the *expected* fitness of the individual, given the frequency with which each prize and delivery time tend to occur in the environment. For any two points in time t and $t+\Delta$, this reduces to comparing the (expected) benefit from the ability to make the distinction between t and $t+\Delta$ against the cost c of supporting the larger brain needed to make the distinction. Denote with $B_t(t,\Delta)$ the expected value at time t of the benefit from distinguishing between t and $t+\Delta$ (a formal derivation of B_t is provided in Appendix A). Nature will provide the individual with the ability to distinguish only if

$$e^{-\delta t} B_t(t,\Delta) \geq c, \tag{1}$$

i.e. only if the present value of the expected benefit outweighs the cost c.

[13] These findings emphasize the dychotomy between the prize and time dimensions when evaluating options, and stand in contrast with an alternative hypotesis, which emphasizes the dychotomy between options involving immediate rewards and those involving delayed rewards (McClure et al., 2004, 2007). Further evidence against that latter hypothesis is found in Kable and Glimcher (2007, 2010). See also the survey article by Kable and Glimcher (2009).

Expression (1) implies that, when both bundles are sufficiently far away in the future, Nature may not want the individual to bear the fitness cost associated to the ability to perceive a difference between them. The reason is that, except for perverse cases, the expected fitness of an individual will normally be bounded.[14] This is sufficient to ensure that B_t is also bounded. As a result, the LHS of (1) approaches zero as t approaches infinity. Hence, there will exist a $t^*(\Delta)$ such that the decision maker does not distinguish between t and $t + \Delta$ for all $t > t^*(\Delta)$. Intuitively, the benefit from distinguishing between time periods gets lower the further these periods are pushed into the future. By contrast, the evolutionary disadvantage incurred takes the form of the individual carrying a bigger, more expensive brain throughout his life, and is therefore independent of the time periods that are being distinguished. Figure 3 compares the fitness benefit at t from distinguishing, $e^{-\delta t} B_t$, to the cost c. The shaded area represents delivery times that are so distant in the future that distinguishing is not optimal.

Consider now prizes. Let $x_1 = x$ and $x_2 = rx$, with $r \geq 0$. Similar to the time dimension, we need to look at the fitness benefit from distinguishing between x and rx. Let then the function $B_x(x, r)$ measure the expected benefit from distinguishing between x and rx (this is formally derived in Appendix A). Nature will endow the individual with the ability to distinguish between x and rx only if

$$B_x(x, r) \geq c. \tag{2}$$

In this case, it will be helpful to normalize all quantities (both costs and

[14] This is obviously the case if $\overline{x}$ is finite. It is clear however that finite expected fitness does not require bounded prizes.

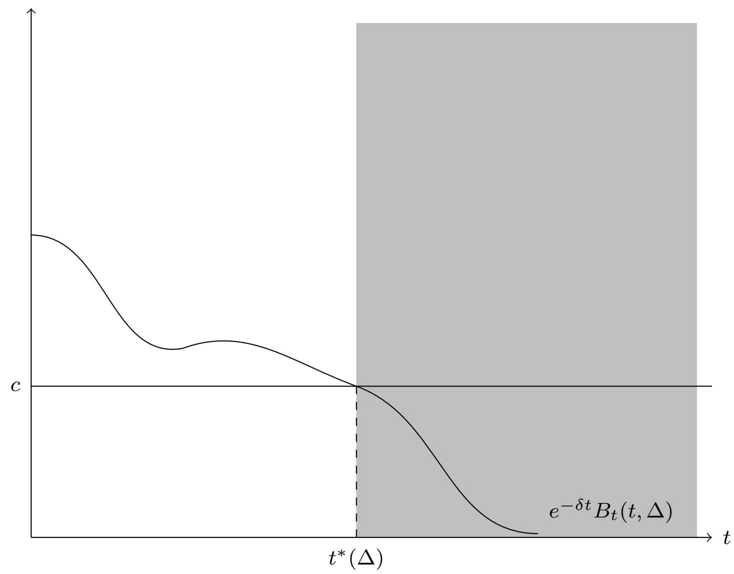

FIGURE 3. COST AND BENEFIT FROM DISTINGUISHING BETWEEN t AND $t+\Delta$.

benefits) by taking their ratio to x.

$$\frac{B_x(x,r)}{x} \geq \frac{c}{x} \tag{3}$$

Notice that, fixing a value for r, the LHS is bounded above.[15] By contrast, c/x becomes arbitrarily large as x approaches zero. It then follows from inequality (3) that, for any r finite, there always exists $x^*(r) > 0$ such that the decision maker does not distinguish between x and rx for all $x < x^*(r)$. Intuitively, if the two prizes are small, the expected benefit from perceiving a difference will necessarily be small as well. Nature will thus not find it worthwhile to endow the individual with the ability to distinguish between the two prizes. Figure 4 compares the fitness benefit relative to x, to the additional cost of sustaining a larger brain relative to x, c/x. Again, the shaded area highlights the set of

[15]Clearly enough, the fitness gain from distinguishing between x and rx cannot exceed $\max\{x, rx\}$. As a result, $B_x(x,r)/x$ cannot exceed $\max\{1, r\}$.

prizes that are so small that they can be safely perceived as similar.

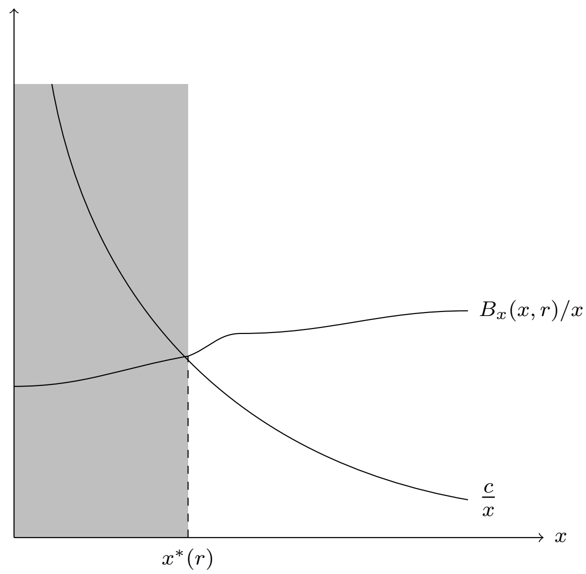

FIGURE 4. COST AND BENEFIT FROM DISTINGUISHING BETWEEN x AND rx.

Of course, the situations depicted in Figures 3 and 4 are not the only possible outcomes. There might be for instance cases where where $e^{-\delta t}B_t$ always lies below c, so that the individual will not distinguish between *any* two points in times whose distance is Δ. This typically happens when Δ is very small. Moreover, $e^{-\delta t}B_t$ and c might cross more than once. Similar arguments apply to B_x/x and c/x.

If we wish to gain better insights into the solution to Nature's problem, we need to be more precise about the benefit side of the problem. A potential difficulty is that the benefit from distinguishing between, say, t and $t + \Delta$ will generally depend on what prizes the individual is able to distinguish. In other words, the solution to Nature's problem for the time dimension is sensitive to the solution for the prize dimension. Symmetrically, the benefit from distinguishing between x and rx will depend on Nature's solution for the time dimension.

It turns out that it is possible to gain valuable insights into the determinants of B_t and B_x by assuming that delivery times are drawn independently of prizes. This allows us to characterize some general properties of B_t and B_x which must hold whatever the solution to Nature's problem in the other dimension.

LEMMA 1: *When delivery times and prizes are mutually independent, the fitness benefit at time t from distinguishing between t and $t + \Delta$ can be decomposed into*

$$B_t(t, \Delta) = g(t, t + \Delta)\phi(\Delta) \tag{4}$$

where ϕ is a continuous function, increasing for $\Delta > 0$, decreasing for $\Delta < 0$, and equal to zero for $\Delta = 0$. Similarly, the fitness benefit from distinguishing between x and rx relative to x can be decomposed into

$$\frac{B_x(x, r)}{x} = h(x, rx)\mu(r) \tag{5}$$

where μ is a continuous function, increasing for $r > 1$, decreasing for $r < 1$, and equal to zero for $r = 1$.

Proof. See Appendix.

The first part of the result essentially says that the function B_t can be broken down into two constituents. The first is the joint probability density g of the two delivery times. The second is a U-shaped function ϕ that only depends on the distance between the two delivery times, Δ, and that reaches a minimum at zero. Let us ignore the density for the moment and focus only on ϕ. The fact that ϕ is U-shaped means that the benefit from distinguishing becomes larger as the distance between the two delivery times moves away from zero. Intuitively, the benefit from distinguishing between two objects is larger when the objects are relatively different than when the objects are relatively similar. This is essentially what the function ϕ captures. The second part of the result says that we can apply a similar reasoning to B_x/x. Ignoring the effect of the density (h) the benefit from distinguishing (μ) becomes larger as the ratio between prizes, r, moves away from one.

Lemma 1 however also shows that the functions B_t and B_x/x will be partly shaped by the distributions of prizes and delivery times. Intuitively, cognitive ability is more beneficial when used to distinguish between pair of objects that occur frequently. The properties of these distributions will therefore contribute to determining whether $t^*(\Delta)$ and $x^*(r)$ are large or small, whether B_t or B_x/x are monotonic in t and x, and how many crossings there will be. The absence of precise empirical restrictions on these distributions adds an additional layer of complexity to the task of rejecting the theory, though. For this reason, we will ignore for the time being the effects of the distributions and analyze a natural benchmark where all bundles are ex-ante equally likely and independent. This helps to show how even a brutally simple (and easily falsifiable) version of the model can go very far in explaining the evidence. Another advantage of analyzing the uniform case is that it makes clear that our results do not rely on different objects occurring with different frequencies. This is in contrast with existing works such as Robson (2001b) and Netzer (2009). Finally, the uniform case allows us to isolate the effects of two rather uncontroversial features of our theory:

1) The fact that the present value of the expected fitness benefit from distinguishing between time periods becomes smaller as the time periods are pushed further away.

2) The fact that the expected fitness benefit from distinguishing between prizes becomes smaller as the prizes become smaller.

We will discuss how the predictions of the model may be affected by the shape of distributions in Section V.

III. Similarity intervals in the uniform model

The precise manner in which the benefits from distinguishing (the functions B_t and B_x) depend on Δ and r determines which delivery times and prizes are deemed by Nature to be sufficiently "close" to be grouped together in the same *similarity set* and thus be perceived as similar by the decision maker. We now provide a characterization of these similarity sets for the benchmark case of uniform prizes and delivery times. In what follows, we thus assume that the delivery time of each bundle is independently drawn from a uniform distribution in $(0, \bar{t})$ and each prize is independently drawn from a uniform distribution in $(0, \bar{x})$. To ensure that the distributions are well defined, both $\bar{t}$ and $\bar{x}$ need to be finite. We will however implicitly assume that they are as large as needed throughout the section.

Given uniform distributions, we can go back to conditions (1) and (3) to compare the benefits from distinguishing with the cost c. Consider the time dimension first. Lemma 1 implies that we can rewrite (1) as

$$e^{-\delta t} \frac{\phi(\Delta)}{\bar{t}^2} \geq c \tag{6}$$

Rather than fixing the difference Δ between delivery times, we can now fix a point in time t and look at the fitness benefit of distinguishing between t and $t + \Delta$ as a function of Δ. This is illustrated in Figure 5. From Lemma 1, we already know that ϕ is U-shaped with a minimum value (of zero) at $\Delta = 0$.

This implies that whenever Δ is between some $\Delta^-(t) < 0$ and $\Delta^+(t) > 0$, (6) will be violated and thus the individual will perceive delivery times t and $t+\Delta$ as similar. The interval $(t+\Delta^-(t), t+\Delta^+(t))$, which contains all delivery times that are perceived as equivalent to time t, is called the *similarity interval* of time t.

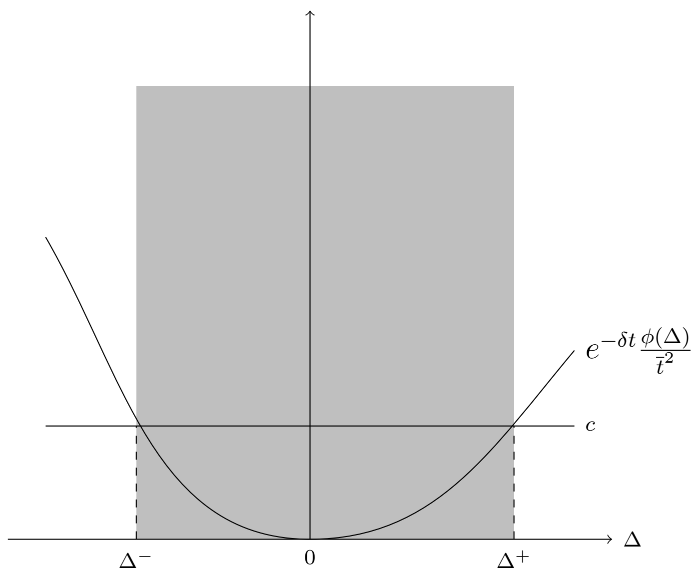

FIGURE 5. SIMILARITY INTERVAL FOR DELIVERY TIMES.

It is possible to determine a similarity interval for any point in time t. Since ϕ does not depend on t, the LHS of (6) is decreasing in t. This implies that larger t will have larger similarity intervals. This is a core insight of our theory: The length of the optimal similarity interval changes with t. As we have seen, the reason is that the present value of the benefit from distinguishing between t and $t+\Delta$ decreases as these time periods are pushed further into the future. By contrast, the fitness cost is independent of the time periods being considered, since it takes the form of the individual carrying a bigger, more expensive brain throughout his life. To develop a graphical intuition of how similarity intervals are determined, it is instructive to rewrite (6) as

$$\frac{\phi(\Delta)}{\bar{t}^2} \geq ce^{\delta t}. \tag{7}$$

Figure 6 depicts the interaction between the LHS of (7) (which is independent of t and is U-shaped in Δ) and its RHS (which is independent of Δ and increasing in t).

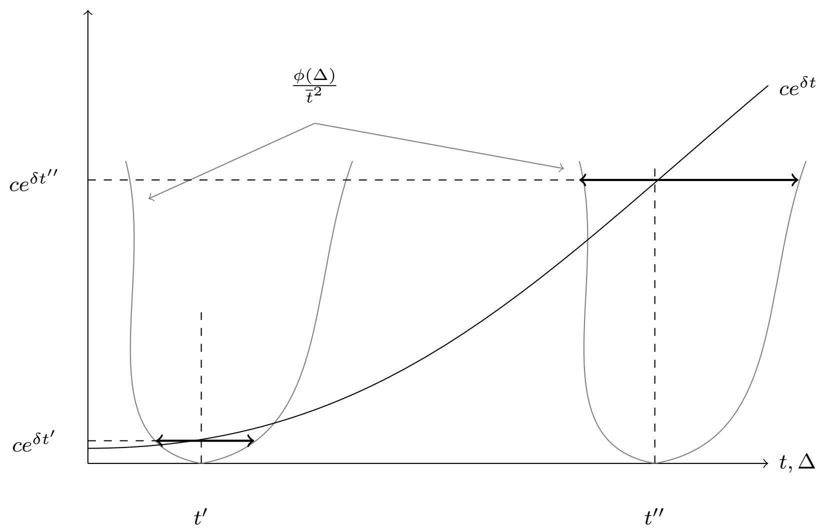

FIGURE 6. LENGTH OF SIMILARITY INTERVALS INCREASES WITH TIME HORIZON.

Consider now prizes. Given Lemma 1, we can rewrite (3) as

$$\frac{\mu(r)}{\bar{x}^2} \geq \frac{c}{x} \tag{8}$$

where $\mu(r)$ is increasing for $r > 1$, decreasing for $r < 1$, and is equal to zero when $r = 1$. Fix a value for x and consider what happens as one changes the ratio r between the two prizes. This is illustrated in Figure 7. For any x, there is always a $r^+(x) > 1$ and a $r^-(x) < 1$ such that, for $r \in (r^+(x), r^-(x))$, (8) will fail to hold, implying that the individual will be unable to distinguish between

x and rx. All prizes whose ratio to x is contained in the interval $(r^-(x), r^+(x))$ are thus perceived as identical to x. We will refer to $(r^-(x), r^+(x))$ as x's similarity interval, keeping in mind that the interval here is an interval of ratios between prizes, rather than an interval of prizes. As in the case of delivery times seen above, the length of a similarity interval depends on x. The RHS of (8) is decreasing in x, implying that larger x have smaller similarity intervals. Again, the intuition is that the gains to be reaped from making fine distinctions between small quantities are not large enough to justify the incurring the cost c of carrying a larger brain.

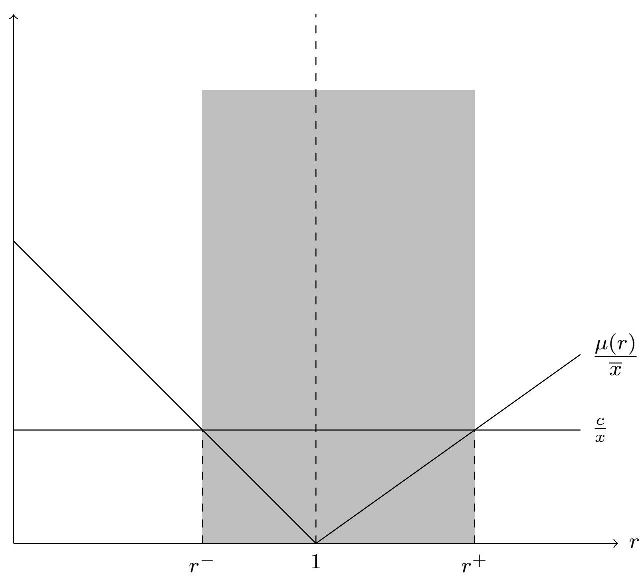

FIGURE 7. SIMILARITY INTERVAL FOR RATIOS BETWEEN PRIZES.

The following Proposition sums up our characterization of the solution to Nature's problem. [In order to keep exposition simple, we abstract from cases where the similarity intervals may partially lie outside the support of t and x.]

PROPOSITION 1: *For every t, Nature chooses a pair $\{\Delta^-(t), \Delta^+(t)\}$, with $\Delta^-(t) < 0 < \Delta^+(t)$, such that the decision maker will not distinguish between t*

and all delivery times $\tau \in (t+\Delta^-(t), t+\Delta^+(t))$*. For every* x*, Nature chooses a pair* $\{r^-(x), r^+(x)\}$*, with* $r^-(x) < 1 < r^+(x)$*, such that the decision maker will not distinguish between* x *and prizes* y *whose ratio to* x *is sufficiently close to one, i.e.* $y/x \in (r^-(x), r^+(x))$*. Moroever,* $\Delta^+(t)$ *is increasing in* t *and* $\Delta^-(t)$ *is decreasing in* t*, so that the length of* t*'s similarity interval increases with* t*. Similarly,* $r^+(x)$ *is decreasing in* x *and* $r^-(x)$ *is increasing in* x*, so that the length of* x'*s similarity interval decreases with* x*.*

Proof. This follows from the graphical argument provided in the text.

We conclude this section by highlighting two properties of similarity intervals. First, they are essentially symmetric.[16] If a is in b's similarity interval, then b will be in a's similarity interval. This follows from the fact that if distinguishing between a and b is suboptimal, it cannot be optimal to distinguish between b and a. On the other hand, they are not consistent with transitivity. If a is perceived similar to b and b to d, similarity between a and d does not generally follow.

IV. Empirical implications of the uniform model

A. Intransitivity of preferences (Cycles)

The available evidence suggests that violations of transitivity in the domain of intertemporal preferences are frequent.[17] These violations are especially challenging theoretically – Manzini and Mariotti (2009) call them a "hard" anomaly. We now show with an example how our setup naturally generates intransitivity. Consider three bundles (x_1, t_2), (x_2, t_2), and (x_3, t_3) with $x_1 <$

[16]The "essentially" here is needed because, when distinguishing and not distinguishing generate exactly the same fitness, we need to assume that Nature consistently chooses the same option (e.g. distinguishing) in order to ensure symmetry.

[17]See, e.g., Tversky et al. (1990) and Roelofsma and Read (2000).

$x_2 < x_3$ and $t_1 < t_2 < t_3$. Assume that the fitness of the three bundles can be ranked in the following way: $x_3 e^{-\delta t_3} < x_1 e^{-\delta t_1} < x_2 e^{-\delta t_2}$. Consider now similarity relationships. Suppose that the differences in prize size between the three bundles are large enough that the individual perceives a difference between all of them. That is, $x_1 \not\approx x_2$, $x_1 \not\approx x_3$ and $x_2 \not\approx x_3$. At the same time, suppose that t_1 is close to the present, whereas t_2 and t_3 are both relatively far away in the future, so that the individual perceives the difference both between t_1 and t_2 and between t_1 and t_3, but perceives t_2 and t_3 as similar. That is, $t_1 \not\approx t_2$, $t_1 \not\approx t_3$ and $t_2 \approx t_3$.

When the individual is faced with the choice between bundles 1 and 2, he will choose (the fitness maximizing) bundle 2. The same applies when he has to choose between 1 and 3: He will go for bundle 1 in this case. On the other hand, when facing a choice between 2 and 3, he will not perceive the difference in delivery times and thus will select bundle 3. As a result, we have a cycle. Bundle 2 is preferred to 1 which is preferred to 3, which is preferred to 2.

B. Time preference reversal (Time Inconsistency)

Consider two bundles (x_1, t_1), (x_2, t_2). In general, we say that a "preference reversal" occurs if the individual initially strictly prefers bundle $i = \{1, 2\}$ but, following a transformation of the two bundles that preserves the fitness ranking, the individual strictly prefers bundle $j \neq i$. Notice that our focus is on the reversal of *preferences*, as opposed to choice reversal. Choice reversal is in principle easier to obtain, for instance by allowing the decision maker to randomize when indifferent between two bundles. We do not consider that, but instead restrict attention to cases of genuine reversal of preferences. The next result deals with time inconsistency.

PROPOSITION 2: *(Time preference reversal) Consider two choices: (i) between* (x_1, t_1) *and* (x_2, t_2)*, with* $t_1 < t_2$ *and* $x_1 < x_2$*, and (ii) between* $(x_1, t_1 + s)$ *and* $(x_2, t_2 + s)$*, for some* $s > 0$*. Then, any reversal of preferences must take the following form: The individual prefers the sooner/smaller bundle* (x_1, t_1) *in (i), and prefers the later/larger bundle* $(x_2, t_2 + s)$ *in (ii). Moreover, for preference reversal to occur, the sooner/smaller bundle must be the fitness maximizing bundle in each of (i) and (ii).*

Proof In Appendix.

Proposition 2 shows that, if a preference reversal occurs, it must take the following form: The individual favors the time dimension when faced with the earlier choice but favors the prize dimension when dealing with the later choice. As seen in the previous Section, Proposition 1 implies that the individual may perceive $t_1 + s$ and $t_2 + s$ as similar and, at the same time, perceive t_1 and t_2 as distinct. Preference reversal then works as follows. When faced with $(x_1, t_1 + s)$ versus $(x_2, t_2 + s)$ – the later choice – the decision maker does not really see any difference between the two time periods, since he perceives them as equally remote. He thus prefers the bundle with the bigger prize. When faced with the earlier choice, by contrast, he *can* see a difference between the two time periods, and is therefore much more inclined to favor earlier consumption over a bigger prize. However, fitness maximization would induce the same behavior in both cases.

As argued in Proposition 2, a necessary condition for time inconsistency is that, in both choices, the sooner/smaller bundle generates higher fitness. This puts an entirely new spin on the phenomenon. Most of the literature implicitly or explicitly assumes that the inefficiencies connected with time inconsistency take the form of early gratification. People, it is said, have an inefficient tendency to favor early consumption. Here, we argue the opposite. People may have an an inefficient tendency to *postpone* consumption, especially when delivery times are far away. Commitment devices aimed at preventing people from grabbing early gratification in favor of later (larger) consumption may thus be efficiency-reducing rather than efficiency-enhancing.[18]

C. Prize preference reversal (Magnitude Effect)

The result in Proposition 1 implies that, fixing the ratio between two prizes, the individual is more likely to distinguish between them when both prizes are large. Intuitively, this is because larger magnitudes create larger gains from making a better choice. Distinguishing between one and two, for instance, generates smaller gains on average than distinguishing between one and two *hundred*. Again, this has implications for preference reversal.

PROPOSITION 3: *(Prize preference reversal) Consider two choices: (i) between* (x_1, t_1) *and* (x_2, t_2)*, with* $x_1 < x_2$ *and* $t_1 < t_2$*, and (ii) between* $(\alpha x_1, t_1)$ *and* $(\alpha x_2, t_2)$ *for some* $\alpha > 1$*. Then, any reversal of preferences must take the following form: The individual prefers the sooner/smaller bundle in (i), and prefers the later/larger bundle in (ii). Moreover, for preference reversal to occur, the later/larger bundle must be the fitness maximizing bundle in each of (i) and (ii).*

Proof In Appendix.

We refer to this form of preference reversal as prize preference reversal or magnitude effect. The individual favors the time dimension when faced with smaller prizes, and favors the prize dimension when faced with bigger prizes. He is less likely to prefer the sooner/smaller option when the prizes involved (in both options) are larger. What really happens is that, when faced with small

[18] These considerations ignore the issue of maladaptation. However, it is possible that conditions in the modern world may be so different from those under which we have evolved that efficiency considerations based on fitness may become void of content.

magnitudes, the decision maker can't see much difference between the two prizes. His preferences over the two bundles are thus shaped by delivery times – he prefers the bundle with less delay. When faced with greater magnitudes, he *can* see a difference between the two prizes, and is therefore much more inclined to postpone consumption in order to obtain a larger prize. However, once again, fitness maximization would induce the same behavior in both cases. Similar to the case of time inconsistency, Proposition 3 identifies which of the two discordant choices observed under prize preference reversal is actually fitness-maximizing. This corresponds to the decision maker's choice when presented with larger magnitudes.

The notion that larger quantities may be discounted less than smaller quantities has been extensively documented.[19] Many empirical studies obtain magnitude effects by eliciting indifference points. For instance, Benhabib et al. (2010) ask questions of the type: "What amount of money y would make you indifferent between x today and y in τ days?", where x is equal to 10/20/30 etc. dollars and τ is equal to 3 days, 2 weeks, 1 month etc. depending on the treatment. In Section B.B1 of the Appendix we show that our setup is consistent with the findings obtained with this type of data.

D. Super/subadditivity (Interval Length Effect)

There is a large body of empirical evidence on time inconsistency and magnitude effects, and the literature treats them almost as stylized facts. Our final observation addresses the issue of interval length effects. Consider for instance the following two choices : (i) between (x, t) and $(rx, t + \Delta)$, where $r > 1$, and (ii) between (x, t) and $(r^k x, t + k\Delta)$ for some $k > 1$. In the first case,

[19]See, e.g., Thaler (1981), Benzion et al. (1989) for early contributions, Benhabib et al. (2010) for a more recent one.

the interval length that separates the two time periods is smaller than in the second case. However, in the second case the prize of the later/larger bundle has been increased to compensate for the greater time delay. The fitness maximizing choice (sooner/smaller or later/larger) is thus the same in both cases. In spite of this, as we will see, the individual will evolve preferences that may lead to inconsistent choices. There are two types of preference reversal that may occur, which we indicate using the following terminology (borrowed from existing literature).

DEFINITION 1: *The decision maker's preferences exhibit subadditivity if he is less patient in (i) (short interval length) than in (ii) (long interval length), i.e.* $(x, t) \succ (rx, t + \Delta)$ *and* $(x, t) \prec (r^k x, t + k\Delta)$*;*
The decision maker's preferences exhibit superadditivity if he is more patient in (i) (short interval length) than in (ii) (long interval length), i.e. $(x, t) \prec (rx, t + \Delta)$ *and* $(x, t) \succ (r^k x, t + k\Delta)$*.*

PROPOSITION 4: *(Interval length effects) Both super and subadditivity may emerge depending on parameter values. A necessary condition for subadditivity is that* $t \not\approx t+\Delta$*. A necessary condition for superadditivity is that* $t \approx t+\Delta$*. In both cases of preference reversal, the decision maker's preferences when faced with the long interval single out the fitness maximizing bundle.*

Proof In Appendix.

The proposition shows that interval length effects may occur in both directions. The individual may exhibit more patience in the choice involving the larger interval (subadditivity), or he may exhibit less patience in that case (superadditivity). There is an empirical literature starting from Read (2001) that documents subadditivity. Subsequent work such as Scholten and Read (2006) also documents the opposite tendency, superadditivity. Moverover, their results suggest that subadditivity is more likely when the short interval is “long,” while superadditivity is more likely when the short interval is “short.” This is in line with our findings. As argued in Proposition 4, a necessary condition for subadditivity is that the short interval must be sufficiently long to ensure that the decision maker perceives the two time periods as distinct. By contrast, a necessary condition for superadditivity is that the short interval must be sufficiently short to render the two time periods indistinguishable. Intuitively, if the decision maker perceives no difference between two time periods, he is unable to appreciate the possible advantage of the smaller/sooner bundle and will therefore never select it. This explains why subadditivity requires that the two time periods in the short horizon case arec perceived as distinct (while the opposite holds for superadditivity).

Note that, for both sub- and superaddivity, the fitness-maximizing bundle corresponds to the decision maker’s preferred bundle when confronted with the long interval treatment. This can serve as a useful guide for normative and policy purposes.

Similar to the case of magnitude effects, many empirical studies obtain interval length effects by eliciting indifference points. These experiments typically work as follows. People are asked to compare (x_1, t_1) with (x_2, t_2); There are two treatments, short interval ($t_2 - t_1$ is small) and long interval ($t_2 - t_1$ is large). The value of x_2 starts low and is progressively increased. The researchers identify the switching point where the decision maker moves from preferring the smaller/sooner to the larger/later bundle, and elicit the implied discount factor. In Section B.B2 of the Appendix we show our setup is consistent with the findings obtained with this type of data.

V. Poisson delivery times and unimodal prize distributions

If we consider delivery times and prizes which are not necessarily equally likely, then Nature's solution will also reflect the frequency with which different bundles occur. Intuitively, Nature will want to allocate scarce cognitive ability to choices that occur more frequently. In this section, we discuss, mostly through examples, what happens in this case. We focus on specific models that are tractable and flexible enough to provide insights on how the shape of the distributions may affect the similarity intervals. In particular, we will assume that delivery times follow a Poisson process and that prizes have a unimodal distribution.

Suppose then that delivery times follow a Poisson process with parameter λ, so that $1/\lambda$ represents the expected waiting time. Let $t = \min\{t_1, t_2\}$ and $\Delta = t_2 - t_1$ so that t is the delivery time of the earlier bundle and $t + |\Delta|$ is the delivery time of the later bundle. Given the nature of the process, t and $|\Delta|$ have independent and identical exponential densities with support $(0, \infty)$. Then, the fitness benefit from distinguishing between the two delivery times can be written as

$$B_t(t, \Delta) = \lambda^2 e^{-\lambda(t+|\Delta|)} \phi(\Delta) \tag{9}$$

where $\lambda^2 e^{-\lambda(t+|\Delta|)}$ is the joint density of t and $t+|\Delta|$. It is accordingly optimal to distinguish between the two delivery times if

$$e^{-(\lambda+\delta)t} \left\{ \lambda^2 e^{-\lambda|\Delta|} \phi(\Delta) \right\} \geq c \tag{10}$$

This is the equivalent of condition (6) in the uniform case. In order to ensure that the LHS of (10) is increasing (decreasing) for $\Delta > 0$ ($\Delta < 0$), we will assume that λ is not too large, i.e. $\lambda < |\phi'(\Delta)|/\phi(\Delta)$ for all Δ.[20] Note that it is as if the term in graphs on the LHS of (6) was discounted using an "adjusted" discount rate $\lambda + \delta$. The usual discount rate δ is modified to take into account the fact that earlier bundles are more likely than late bundles and, thus, from an ex-ante viewpoint, are more "important."

Consider now what happens when, by changing λ, we affect the expected waiting time of both bundles. For this purpose, it is instructive to rewrite (10) as

$$\lambda^2 e^{-\lambda|\Delta|}\phi(\Delta) \geq ce^{(\lambda+\delta)t} \tag{11}$$

An increase in λ makes the RHS of (11) steeper in t. This implies that the length of the similarity intervals becomes more sensitive to time. Intuitively, short expected waiting times imply that, from Nature's ex-ante perspective, the bulk of the individual's cognitive capabilities should focus on distinguishing between short run bundles, since these are those which are likely to occur. This leads to larger similarity intervals for larger t, as shown in Figure 8. However, an increase in λ also affects the LHS of (11). This becomes steeper for values of Δ close to zero and flatter for values of Δ away from zero. Intuitively, shorter expected waiting times imply that the time interval between the two bundles is likely to be small. As a result, Nature wants to allocate more cognitive power to distinguishing between bundles that are relatively similar in the time dimension.

While the Poisson model is elegant and tractable, it is clear that the core

[20]If the expected waiting times are very short Nature may not endow the individual with the ability to distinguish between any two points in time. Intuitively, if both bundles always occur in the present or in the very short run, there is little value in considering the time dimension at all.

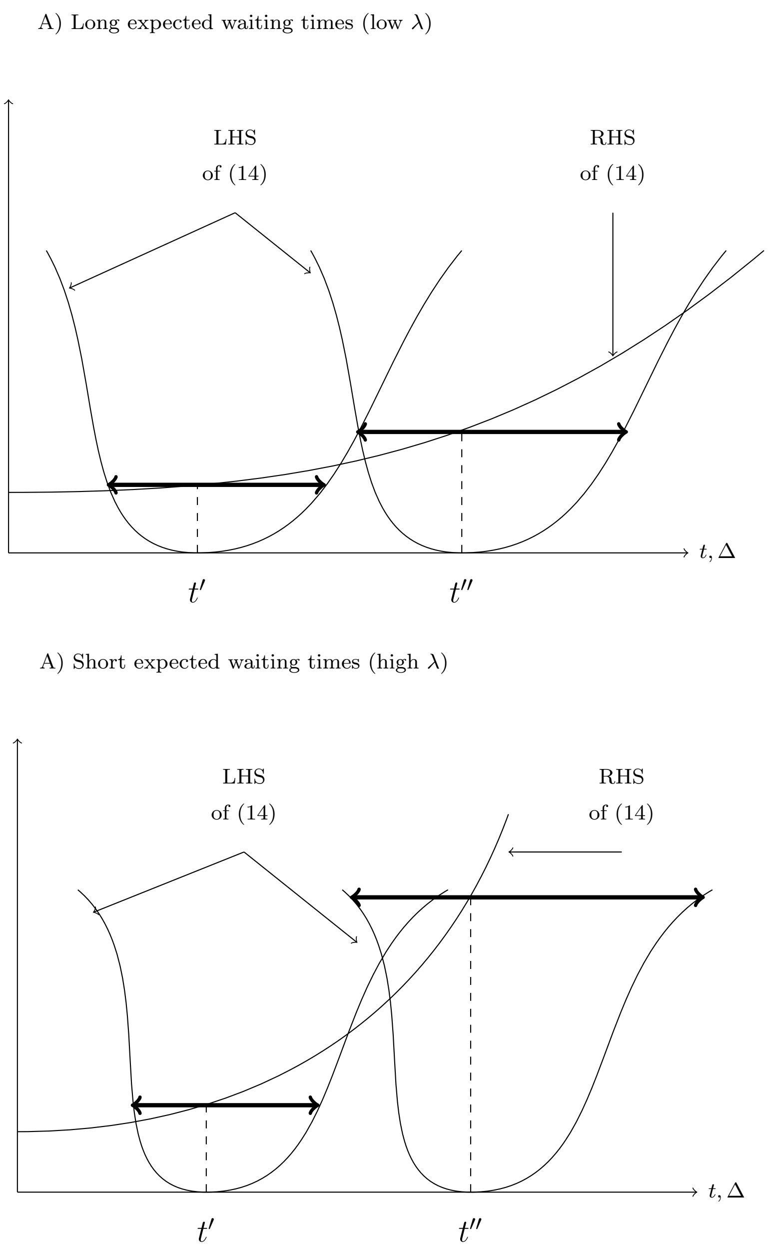

Figure 8. Similarity intervals for Poisson processes with long and short expected waiting times.

results would apply to any process where delivery times have a joint density that is weakly decreasing in t. Consider now quantities. Suppose that, similar to the case of Poisson delivery times, the ratio r between the two quantities is statistically independent of the magnitude of x. In other words, the chances that a \$20 prize is drawn given a \$10 prize is the same as the chance that a \$2,000 prize is drawn given a \$1,000 prize. Formally, let $x \equiv \max\{x_1, x_2\}$ and $r \equiv x_1/x_2$ and assume that the joint density $h(x_1, x_2)$ can be written as

$$h = h_r(|\ln r|)h_x(x) \tag{12}$$

where h_r and h_x are continuous density functions with full support in $(0, \infty)$.[21] It is then optimal to distinguish between x and rx if

$$h_r(|\ln r|)\mu(r) \geq c/xh_x(x). \tag{14}$$

This is the equivalent of inequality (8) in the uniform case. The RHS represents the cost relative to x, which has been "adjusted" to take into account the fact that some quantities occur more frequently than others. This is obtained by weighting x by its probability density $h_x(x)$. It is particularly interesting to look at the case where $h_x(x)$ has a unique interior maximum, so that very small and very large quantities tend to occur infrequently. If $h_x(x)$ has a unique interior maximum, the RHS of (14) may have a global minimum. This is illustrated in Figure 9 (where we took for simplicity an improper uniform for h_r).

[21] Specifying h_r in terms of the absolute value of the log of r ensures that x_1 and x_2 are exchangeable, i.e.

$$h(x_1, x_2) = h_r(|\ln x_1 - \ln x_2|)h_x(\max\{x_1, x_2\}) = h(x_2, x_1) \text{ for all } (x_1, x_2) \tag{13}$$

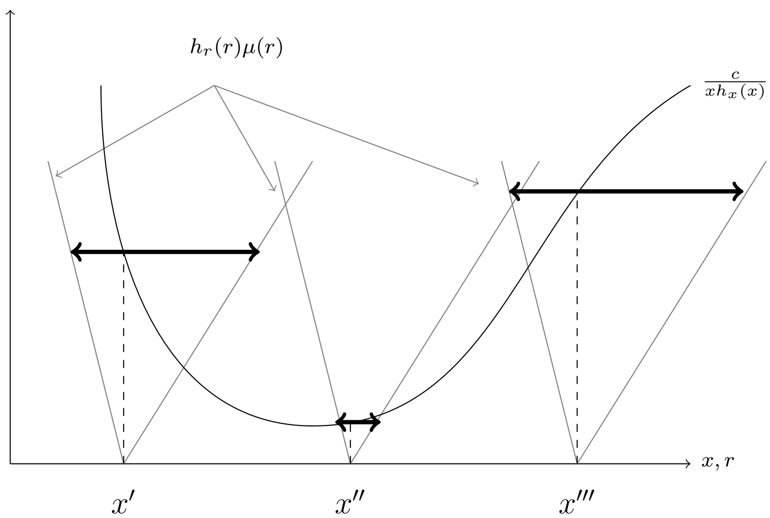

FIGURE 9. SIMILARITY INTERVALS WHEN VERY SMALL AND VERY LARGE QUANTITIES ARE INFREQUENT.

Notice that the length of the similarity intervals will generally be non-monotonic in x. Intuitively, for small values of x (e.g. $x = x'$ in Figure 9) both the fitness benefit from distinguishing and the frequency of the two prizes are small. Hence, Nature will choose to endow the individual with the ability to distinguish between two quantities only if their ratio r is quite far away from one. For values of x close to the modal value (e.g. $x = x''$), both the large frequency and the larger fitness benefits from distinguishing make it worthwhile to endow the individual with the ability to make relatively fine distinctions. However, when x is very large ($x = x'''$) the fitness gain from distinguishing is large but the frequency is small. If the second effect dominates the first (i.e. very large bundles are extremely rare), then Nature will provide the individual with the ability to distinguish between very large bundles only when their ratio is relatively far away from one. Indeed, introspection suggests that we are equally bad at perceiving the difference between 1¢ and 2¢ and the difference between 10 and 20 billions. On the other hand, most people have no problem in perceiving the difference between \$100 and \$200. Notice however that for low and moderate stakes, the size of the similarity interval is decreasing in the value of x. As a result, a magnitude effect would still be present when stakes change from low to moderate.

VI. Related literature

The purpose of our analysis is to build an evolutionary model of intertemporal choice. Our approach may be seen either as providing an entirely novel account, or as proposing a possible *microfoundation* of descriptive accounts, thus complementing them. In this light, it is instructive to compare more closely our predictions with those of existing work.

A. Descriptive models of time preferences

In this section we review some existing axiomatic/descriptive models of time preferences and we discuss their predictions with respect to the phenomena we have identified, namely cycles, time inconsistency, magnitude effect and interval length effects, in the forms we have described in the previous section.[22] In what follows, we keep with the literature and focus on linear (consumption) utilities. As we will see, the restriction to linearity makes many of the models incompatible with the magnitude effect.[23]

i) Exponential discounting Exponential discounting (Samuelson, 1937) is no-

[22]We can only provide a brief summary here. Manzini's and Mariotti's (2009) extensive literature review presents a taxonomy of each model's predictions in terms of cycles, time inconsistency and magnitude effects. In a Section B of the Appendix, we also provide proofs of our statements about each model's ability to predict interval length effects. Rubinstein (2003) and Read (2001) are omitted since, in spite of making important contributions, they do not propose fully fledged models. Ok and Masatlioglu (2007) derive a tractable mathematical format that covers a wide variety of descriptive models of time preferences.

[23]However, sufficient concavity of the utility function would restore compatibility with the magnitude effect. The question then becomes whether such concavity is or is not justifiable (see Rabin, 2000).

toriously incompatible with time inconsistency, cycles, and with interval length effects. If we restrict attention to linear utility, it is also incompatible with the magnitude effect.

ii) Hyperbolic discounting A number of researchers have posited a specific functional form, hyperbolic discounting, to account for observed tendencies for immediate gratification – see e.g., Ainsle (1991), Loewenstein and Prelec (1992) and Laibson (1997). Hyperbolic discounting is compatible with time inconsistency and interval length effects, but it is incompatible with cycles. If we restrict attention to linear utility, it is also incompatible with the magnitude effect.

iii) Manzini and Mariotti's (2006) (σ, δ)*-model* Manzini and Mariotti (2006) propose a theory of "vague time preferences" (VTP). This theory focuses on the individual's ability to compare between different bundles. In their model, utility is discounted exponentially. However, if two bundles generate utility levels that are sufficiently close, then the agent is unable to choose based on utility and resorts to some additional heuristic to make his choice. The model can thus be thought of as describing the possible implications of the agent's inability to distinguish between similar shades of utility. Here, we concentrate on the simplest specification of VTP, namely the (σ, δ)-model. If the two bundles generate utility levels that differ by more than a fixed amount σ, then the option yielding the highest utility is preferred. If the difference between utilities is less than σ, then the individual will prefer the alternative with the earlier delivery time (in the time prominence variant of the model) or the greater prize (in the outcome prominence variant of the model). Manzini and Mariotti (2006) focus on the outcome prominence version of the (σ, δ)-model, since the time prominence variant is incompatible with time inconsistency.[24]

[24]Details are provided in Section B.B3 of the Appendix.

If we restrict attention to linear utility, the (outcome prominence version of) the (σ, δ) model is incompatible with the magnitude effect.[25] The model is compatible with cycles, and with interval length effects, but only in the form of superadditivity.[26]

iv) *Benhabib et al.'s (2010) fixed cost model* In Benhabib et al.'s (2010) model, individual preferences are as follows: $u(x, 0) = x$, $u(x, t) = \rho^t x - b$ $\forall t > 0$, for some strictly positive constant b and some discount factor ρ. This fixed cost model is compatible with time inconsistency, with the magnitude effect, and with subadditivity (but not superadditivity).[27] However, it can explain these preference reversals only when the earliest alternative occurs at time 0. It is also incompatible with cycles.

B. Other related literature

The contribution of our paper is to supply an "ultimate" cause that complements "proximate" descriptions. We are not aware of any work that provides a rationale for the whole range of phenomena we are able to account for. The literature on the evolutionary foundations of preferences has concentrated primarily of the phenomenon of time preference reversal. The two key references here are Dasgupta and Maskin (2006) and Netzer (2009).

Dasgupta and Maskin (2006) consider a setup where there is uncertainty about the timing when a chance to consume will occur. Preference reversal emerges because the available information changes over time. As a result, the baseline model produces no dynamic inconsistency: Dynamic inconsistency may only emerge from a mismatch between the environment in which preferences were formed and the current environment. By contrast, our setup

[25] However, it would be if we allowed the parameter σ to depend on magnitude.
[26] A proof of this last statement is provided in Section B.B3 of the Appendix.
[27] A proof of this last statement can be found in Section B.B4 of the Appendix.

generates dynamic inconsistency even in the absence of mismatch.[28]

Netzer (2009) studies the evolution of preferences in a setup where individuals face a choice either between two short-run alternatives (i.e., $t_1 = t_2 = 0$) or between two long-run alternatives (i.e., $t_1 = 1$ and $t_2 = 2$). Nature can activate different decision mechanisms depending on the type of decision (short-run or long-run). If short-run decisions and long run decisions are characterized by different frequencies of small payoffs (relative to other payoffs), the optimal mechanisms may generate what the author calls "regret": The individual selects the alternative with the longer waiting time in the long-run decision, but would like to revert this choice after one period. Dynamic inconsistency thus emerges (and takes the form of regret). However, dynamically inconsistent behavior cannot arise within the model, or, if it does, it must be the result of maladaptation. By contrast, our setup may well generate dynamically inconsistent behavior.[29] Moreover, as we have shown, our explanation is quite different, since it does not require the distribution of prizes to change with time in a particular way (see Section III).[30] By contrast, the gist of the argument in Netzer (2009) relies on the properties of the distribution of payoffs.

Robson and Samuelson (2009) consider the evolution of discount rates in the presence of aggregate shocks, where the effects of aggregate shocks may differ across ages. Under some natural conditions, they find that the discount

[28]To see how the preferences of, say, the time-0 incarnation of an individual may differ from those of his time-t incarnation, consider the comparison between (x_1, t_1) and (x_2, t_2), where $x_2 > x_2$ and $t_2 > t_1$. At $t = 0$, the decision maker may prefer the later-larger bundle (since he doesn't perceive any difference between t_1 and t_2). At a later date, however, the agent will eventually perceive a difference between t_1 and t_2, and may thus prefer the sooner-smaller bundle.

[29]In fact, in our setup the possibility of reverting one's earlier decision at a later date would improve the quality of decision-making (this is a direct consequence of the observation that, whenever time preference reversal occurs the "myopic" self takes the better decision in terms of fitness).

[30]However, as discussed in section V, this does not imply that probability distributions are unimportant in our analysis.

rate will fall with age. Thus, evolution exhibits a "present bias," but this bias does not lead to preference reversals, since the discount rates change with *age*, rather than with *time*.[31]

Our work is also related to Kőszegi and Szeidl (2013), who consider a setup where agents maximize a "focus-weighted" utility. People are assumed to focus more on (and hence overweight) attributes in which their options differ more. This generates a present bias and time inconsistency whenever the future effect of a current decision is distributed over many dates – for instance, a person may focus too little on the small future health benefits of exercise relative to the big current cost. Clearly, this differs from our story.

Finally, our results exhibit parallels with the psychology theory of *diminishing sensitivity* (Tversky and Fox 1995, Tversky and Kahneman 1992), which states that people are more sensitive to changes near their status quo (typically, the moment when the decision is taken, i.e. "now") than to changes remote from their status quo. According to this theory, people feel more of a difference between, say, now and four weeks than between 26 and 30 weeks from now. Our setup provides a possible explanation, which relies on the idea that the gains to be reaped from endowing the agent with the ability to make the first distinction are large – since they arise relatively soon – while the gains from making the second distinction are small – since they are far in the future. This provides a powerful intuition for time preference reversal.

[31]Robson and Samuelson (2007) also obtain that the rate of discounting should fall with age.

VII. Concluding remarks

We hope that our work may illustrate the usefulness of looking at "ultimate" causes to complement descriptive or "proximate" accounts. In their survey article, Robson and Samuelson (2011) ask the following question:

> Must evolutionary models of preference reversals necessarily involve mismatches, or are there circumstances under which evolutionary design calls for preference reversals in the environment giving rise to that design?

In contrast with previous literature, the model we have presented provides an instance of the latter. People don't distinguish between faraway periods and may thus end up taking suboptimal decisions. However, this is part of Nature's optimal solution when designing similarity intervals.

Our analysis also adds to existing work by casting light on the welfare implications of preference reversals. For instance, we find that, in the presence of time inconsistency, the "impulsive" self makes the right (fitness-maximizing) choice, while the "far-sighted" self is too patient. This is at odds with many current interpretations of this phenomenon.

Finally, we conjecture that our approach may also prove useful to the study of choice under risk. For instance, a similar rationale to the one given for our magnitude effect may explain why individuals perceive a difference between a probability of 1 and one of 0.8, while they perceive no difference between a probability of 0.25 and one of 0.2, despite the ratio between the two numbers being the same in both cases. As observed by Rubinstein (1988), this may help shed light on puzzling phenomena, such as the Allais paxadox. We believe that this may provide a fruitful avenue for future research.

REFERENCES

Aiello, L.C., and Wheeler, P. (1995). The expensive-tissue hypothesis: the brain and the digestive system in human and primate evolution, *Current Anthropology* 36: 199–221.

Ainsle, G. (1991) Derivation of 'rational' economic behavior from hyperbolic discount curves, *American Economic Review*, 81: 334-40.

Alger, I., and Weibull, J.B., (2013) Homo moralis – Preference evolution under incomplete information and assortative matching. *Econometrica* 81: 2269-2302.

Ballard, K.and Knutson, B. (2009) Dissociable neural representations of future reward magnitude and dealy during temporal discounting, *Neuroimage* 45:143-150.

Benhabib, J., Bisin, A. and Schotter, A. (2010) Present bias, quasi-hyperbolic discounting and fixed costs, *Games and Economic Behavior*, 69: 205–223.

Benzion, U., Rapoport, A. and Yagil, J. (1989) Discount rates inferred from decisions: An experimental study, *Management Science* 35: 270-284.

Bernheim, B. D. (2009) Behavioral welfare economics, *Journal of the European Economic Association*, 7: 267-319.

Bernheim, B. D., and Rangel, A. (2009) Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics, *Quarterly Journal of Economics* 124: 51-104.

Binmore, K. (2005) Natural Justice, Oxford: Oxford University Press.

Dasgupta P. and Maskin E. (2006) Uncertainty and hyperbolic discounting, *American Economic Review*, 95:1290–1299.

Frank, R. H. (1987) If homo economicus could choose his own utility function, would he want one with a conscience? *American Economic Review* 77: 593-604.

Frederick, S., Loewenstein, C., and O'Donoghue, T. (2002) Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40: 351-401.

Fudenberg, D., and D. K. Levine (2006) A dual-self model of impulse control, *American Economic Review* 96: 1449-1476.

Hansson, I., and Stuart, C. (1990) Malthusian selection of preferences, *American Economic Review* 80: 529–544.

Kable, J. W. and Glimcher, P. W. (2007) The neural correlates of subjective value during intertemporal choice, *Natural Neuroscience* 10: 1625-1633.

Kable, J. W. and Glimcher, P. W. (2009) The neurobiology of decision: Consensus and controversy, *Neuron* 63: 733-745.

Kable, J. W. and Glimcher, P. W. (2010) The "As Soon As Possible" effect in human intertemporal decision making: Behavioral evidence and neural mechanisms, *Journal of Neurophysiology* 103: 2513-2531.

Kőszegi, B. and Szeidl, A. (2013) A model of focusing in economic choice, *Quarterly Journal of Economics* 128: 53-104.

Laibson, D. (1997) Golden eggs and hyperbolic discounting, *Quarterly Journal of Economics* 112: 443-477.

Leland, J. W. (2002) Similarity judgements and anomalies in intertemporal choice, *Economic Inquiry* 40: 574-581.

Loewenstein, G. and Prelec, D. (1992) Anomalies in intertemporal choice: evidence and interpretation, *Quarterly Journal of Economics* 107: 573-597.

Lovejoy, C. O. (1088) Evolution of human walking, *Scientific American* 259.5: 82-89.

Manzini, P. and Mariotti, M. (2006) A vague theory of choice over time, *Advances in Theoretical Economics* 6: article 6.

Manzini, P. and Mariotti, M. (2009) Choice over time, in: P. Anand, P. Pattanaik and C. Puppe (eds.) *Oxford Handbook of Rational and Social Choice*, Oxford.

McClure, S. M., Laibson, D. I., Loewenstein, G., Cohen, J. D. (2004) Separate neural systems value immediate and delayed monetary rewards, *Science* 306: 503-507.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., Cohen, J. D. (2007) Time discounting for primary rewards, *Journal of Neuroscience*, 27: 5796-5804.

Netzer, N. (2009) Evolution of time preferences and attitudes toward risk, *American Economic Review*, 99: 937-955.

O'Donoghue, T., and Rabin, M. (1999) Doing it now or later. *American Economic Review*, 89: 103-124.

O'Donoghue, T., and Rabin, M. (2003) Studying optimal paternalism, illustrated by a model of sin taxes. *American Economic Review* 93: 186-191.

O'Donoghue, T., and Rabin, M. (2006) Optimal sin taxes. *Journal of Public Economics*, 90: 1825-1849.

Ok, E. A. and Y. Masatlioglu (2007) A theory of (relative) discounting, *Journal of Economic Theory* 137: 214-245.

Phelps, E. S. and Pollak, R. A. (1968) On second–best national saving and game–equilibrium growth. *Review of Economic Studies*, 35: pp. 185-99.

Pine, A., Seymour, B., Roiser, J. P., Bossaerts, P., Friston, K. J., Curran, H. V., Dolan, R. J. (2009) Encoding of marginal utility across time in the human brain, *Journal of Neuroscience*, 29: 9575-9581.

Rabin, M. (2000). Risk aversion and expected utility theory: A calibration theorem. *Econometrica*, 68: 1281-1292.

Read, D. (2001), Is time–discounting hyperbolic or subadditive? *Journal of Risk and Uncertainty*, 23: 5-32.

Robson, A. J. (2001a) Why would nature give individuals utility functions? *Journal of Political Economy*, 109: 900-914.

Robson, A. J. (2001b) The biological basis of economic behavior, *Journal of Economic Literature*, 39: 11-33.

Robson, A.J., and L. Samuelson (2007) The evolution of intertemporal preferences, *American Economic Review* 97: 496–500.

Robson, A. J. and Samuelson, L. (2009) The evolution of time preference with aggregate uncertainty, *American Economic Review* 99: 1925-1953.

Robson, A. J. and Samuelson, L. (2011a) The evolutionary foundations of preferences, in: A. Bisin, M. O. Jackson (eds.) *Handbook of Social Economics*, North-Holland.

Robson, A. J. and Samuelson, L. (2011b) The evolution of decision and experienced utilities, *Theoretical Economics* 6: 311-339.

Roelofsma, P. H. and Read, D. (2000) Intransitive intertemporal choice, *Journal of Behavioral Decsion Making*, 15: 433-460.

Rogers A. R. (1994) Evolution of time preferences by natural selection, *American Economic Review* 84: 460-481.

Rubinstein, A. (1988) Similarity and decision-making under risk (Is there a utility theory resolution to the Allais paradox?), *Journal of Economic Theory* 46: 145-153.

Rubinstein, A. (2003) Is it "Economics and psychology"? The case of hyperbolic discounting, *International Economic Review* 44: 1207-1216.

Samuelson, L. (2004) Information-based relative consumption effects, *Econometrica* 72: 93-118.

Samuelson, P.A. (1937) A note on measurement of utility, *Review of Economic Studies*, 4: 155-161.

Samuelson, L. and Swinkels, J. (2006) Information, evolution and utility, *Theoretical Economics* 1: 119-142.

Scholten, M. and Read, D. (2006) Discounting by intervals: A generalized model of intertemporal choice, *Management Science* 52(9):1424-1436.

Strotz, R. H. (1956) Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23: 165-80.

Thaler, R. (1981) Some empirical evidence on dynamic inconsistency, *Economics Letters* 8: 201-207.

Tversky, A., Slovic, P. and Kahneman, D. (1990) The causes of preference reversal, *American Economic Review* 80: 204-217.

Tversky, A. and Kahneman, D. (1992) Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty* 5: 297-323.

Tversky, A. and Fox, C. F. (1995) Weighing risk and uncertainty, *Psychological Review* 102: 269–283.

Waldman, M. (1994) Systematic errors and the theory of natural selection, *American Economic Review*, 84: 482-49.

Mathematical Appendix (for online publication)

A1. Notation and preliminaries

Nature's problem essentially reduces to finding a similarity set for each t and a similarity set for each x such that overall expected fitness is maximized. More formally, we denote with

$$S(t) = \{\tau \in (0, \bar{t}) : \tau \approx t\}. \tag{A1}$$

the optimal similarity correspondence for delivery times $t \in (0, \bar{t})$. In the same way, we denote with

$$\Sigma(x) = \{y \in (0, \overline{x}) : y \approx x\}. \tag{A2}$$

the optimal similarity correspondence for prizes $x \in (0, \overline{x})$. The complementary correspondences are denoted with $S_c(t)$ and $\Sigma_c(x)$ (i.e. $S_c(t) = \{\tau \in (0, \bar{t}) : \tau \not\approx t\}$ and $\Sigma_c(x) = \{y \in (0, \overline{x}) : y \not\approx x\}$).

It is clear that any solution to Nature's problem in the time dimension must be such that, for any t and Δ,

$$e^{-\delta t} B_t(t, \Delta) < c \Rightarrow t + \Delta \in S(t) \tag{A3}$$

$$e^{-\delta t} B_t(t, \Delta) > c \Rightarrow t + \Delta \in S_c(t), \tag{A4}$$

where the function B_t denotes the expected benefit at time t from distinguishing between t and $t + \Delta$. Symmetrically, for the prize dimension,

$$B_x(x, r) < c \Rightarrow rx \in \Sigma(x) \tag{A5}$$

$$B_x(x, r) > c \Rightarrow rx \in \Sigma_c(x), \tag{A6}$$

where B_x is the expected benefit at time zero from distinguishing between x and rx.

To characterize the solution to Nature's problem, we thus need to provide an explicit derivation of the expected benefit functions B_t and B_x. The main problem is that B_t will generally depend on the shape of $\Sigma(x)$ and B_x on the shape of $S(t)$. It is however possible to characterize a number of properties of B_t (resp. B_x) that do not depend on the precise shape of $\Sigma(x)$ (resp. $S(t)$). [In order to avoid unnecessary complications, we will restrict attention to the case where, fixed any x (resp. t), the resulting similarity set $\Sigma(x)$ (resp. $S(t)$) is an interval or a collection of disjoint intervals of positive length.]

A2. Derivation of B_t and B_x and symmetry of similarity relations

In this section we derive B_t and B_x and establish symmetry of similarity relations. Let $t_1 = t$ and $t_2 = t + \Delta$. We will first consider the case $\Delta > 0$ and then apply a symmetry argument to the case $\Delta < 0$. Denoting with x_1 the prize associated with t and with x_2 the prize associated to $t + \Delta$, the actual fitness at time t of an individual who perceives the difference between t and $t + \Delta$ is as follows.

- If $x_1 \not\approx x_2$, time t fitness is

$$\text{(A7)} \qquad \begin{cases} x_1 & \text{if } x_2 \in [0, x_1 e^{\delta\Delta}] \\ e^{-\delta\Delta} x_2 & \text{if } x_2 \in (x_1 e^{\delta\Delta}, \overline{x}) \end{cases}$$

- If $x_1 \approx x_2$, time t fitness is equal to x_1.

Denote with $p_{t,\Delta}(x_1, x_2)$ the joint density of the pair (x_1, x_2) conditional on t being the delivery time of x_1 and $t + \Delta$ being the delivery time of x_2. Conditional on t and $t + \Delta$ being the arrival times of the two bundles, the ex-ante expected fitness for a decision maker who perceives the difference between t and $t + \Delta$ is

$$\text{(A8)} \qquad \int_{x_1} \left[\int_{x_2 \in \Sigma_c(x_1) \cap [0, x_1 e^{\delta\Delta}]} x_1 p_{t,\Delta}(x_1, x_2) dx_2 + \int_{x_2 \in \Sigma_c(x_1) \cap (x_1 e^{\delta\Delta}, \overline{x})} e^{-\delta\Delta} x_2 p_{t,\Delta}(x_1, x_2) dx_2 + \int_{x_2 \in \Sigma(x_1)} x_1 p_{t,\Delta}(x_1, x_2) dx_2 \right] dx_1$$

Consider now the same for a decision maker who does not perceive the difference between t and $t+\Delta$,

- If $x_1 \not\approx x_2$, time t fitness is

$$
\text{(A9)} \qquad \begin{cases} x_1 & \text{if } x_2 \in [0, x_1] \\ e^{-\delta\Delta} x_2 & \text{if } x_2 \in (x_1, \overline{x}) \end{cases}
$$

- If $x_1 \approx x_2$, time t fitness is on average $\frac{1}{2}(x_1 + e^{-\delta\Delta}x_2)$.

Hence, conditional on t and $t+\Delta$ being the arrival times of the two bundles, the ex-ante expected fitness at time t is

$$
\text{(A10)} \qquad \int_{x_1} \left[\int_{x_2 \in \Sigma_c(x_1) \cap [0, x_1]} x_1 p_{t,\Delta}(x_1, x_2) dx_2 + \int_{x_2 \in \Sigma_c(x_1) \cap (x_1, \overline{x})} e^{-\delta\Delta} x_2 p_{t,\Delta}(x_1, x_2) dx_2 + \int_{x_2 \in \Sigma(x_1)} \frac{1}{2}(x_1 + e^{-\delta\Delta} x_2) p_{t,\Delta}(x_1, x_2) dx_2 \right] dx_1
$$

The expected benefit at time t from distinguishing between t and $t+\Delta$ is obtained by multiplying the difference between (A8) and (A10) times the joint density of t and $t+\Delta$ (denoted by g)

$$
\text{(A11)} \qquad B_t(t, \Delta) = g(t, t+\Delta) \int_{x_1} \left[\int_{x_2 \in \Sigma_c(x_1) \cap [x_1, x_1 e^{\delta\Delta}]} (x_1 - e^{-\delta\Delta} x_2) p_{t,\Delta}(x_1, x_2) dx_2 + \int_{x_2 \in \Sigma(x_1)} \frac{1}{2}(x_1 - e^{-\delta\Delta} x_2) p_{t,\Delta}(x_1, x_2) dx_2 \right] dx_1
$$

Notice now that, trivially, the present value of the expected benefit from distinguishing between any t and $t+\Delta$ must be the equal to the present value of the expected benefit from distinguishing between $t+\Delta$ and t. This implies that the identity

$$
\text{(A12)} \qquad e^{-\delta(t+\Delta)} B_{t+\Delta}(t+\Delta, -\Delta) \equiv e^{-\delta t} B_t(t, \Delta)
$$

holds for any t and Δ. Using (A11) and (A12) it is then immediate to determine $B_t(t,\Delta)$ for $\Delta < 0$. Moreover, from (A12),

$$e^{-\delta(t+\Delta)}B_{t+\Delta}(t+\Delta,-\Delta) \gtreqless c \Leftrightarrow e^{-\delta t}B_t(t,\Delta) \gtreqless c. \quad \text{(A13)}$$

This ensures that $t+\Delta \in S(t) \Leftrightarrow t \in S(t+\Delta)$, which in turn implies that similarity relations are symmetric. Finally, notice that, for all t, $B_t(t,0) = 0$, i.e. the expected benefit at time t from distinguishing between t and t itself must be zero. [Notice however that, for a generic conditional density $p_{t,\Delta}$, the limit of (A11) for $\Delta \to 0$ need not be zero. As we show below, for continuity to hold at $\Delta = 0$, x_1 and x_2 need to be exchangeable. We will come back to this point in the proof of Lemma 1.]

We now repeat the same exercise for prizes. Let $x_1 = x$ and $x_2 = rx$. We start off with $r > 1$ and then analyze the case $r < 1$. Consider then the actual fitness of a decision maker who perceives the difference between x and rx. Denoting with t_1 the delivery time associated with x and with t_2 the delivery time associated with rx, it is clear that, a) if $t_1 \not\approx t_2$, fitness is equal to $rxe^{-\delta t_2}$ if $t_2 \in [0, t_1 + \ln r/\delta]$ and is equal to $xe^{-\delta t_1}$ if $t_2 \in (t_1 + \ln r/\delta, \bar{t})$; b) if $t_1 \approx t_2$, then fitness is always equal to $rxe^{-\delta t_2}$. Hence, conditional on prizes being x and rx, the expected fitness of a decision maker who perceives a difference between x and rx is

$$x\int_{t_1}\Bigg[\int_{t_2\in S_c(t_1)\cap[0,t_1+\ln r/\delta]} re^{-\delta t_2}p_{x,r}(t_1,t_2)dt_2 + e^{-\delta t_1}\int_{t_2\in S_c(t_1)\cap(t_1+\ln r/\delta,\bar{t})} p_{x,r}(t_1,t_2)dt_2 + \int_{t_2\in S(t_1)} re^{-\delta t_2}p_{x,r}(t_1,t_2)dt_2\Bigg]dt_1 \quad \text{(A14)}$$

where $p_{x,r}(t_1,t_2)$ is the joint density of the pair (t_1,t_2) conditional on x and r. Consider now a decision maker who does not distinguish between x and rx. Then, a) if $t_1 \not\approx t_2$, his fitness will be equal to $rxe^{-\delta t_2}$ if $t_2 \in [0, t_1]$ and will be equal to $xe^{-\delta t_1}$ if $t_2 \in (t_1, \bar{t})$; b) if $t_1 \approx t_2$, his fitness will be on average $\frac{1}{2}(xe^{-\delta t_1} + rxe^{-\delta t_2})$. As a result, conditional on prizes x and rx, his expected fitness is

$$x\int_{t_1}\Bigg[\int_{t_2\in S_c(t_1)\cap[0,t_1]} re^{-\delta t_2}p_{x,r}(t_1,t_2)dt_2 + e^{-\delta t_1}\int_{t_2\in S_c(t_1)\cap(t_1,\bar{t}]} p_{x,r}(t_1,t_2)dt_2 + \int_{t_2\in S(t_1)}\frac{1}{2}[e^{-\delta t_1}+re^{-\delta t_2}]p_{x,r}(t_1,t_2)dt_2\Bigg]dt_1 \tag{A15}$$

The ex-ante fitness benefit of distinguishing between x and rx is given by the difference between (A.A2) and (A15) times the joint density of x and rx. Expressing this quantity as a percentage of x, we obtain

$$\frac{B_x(x,r)}{x} = h(x,rx)\int_{t_1}\Bigg[\int_{t_2\in S_c(t_1)\cap[t_1,t_1+\ln r/\delta]}(re^{-\delta t_2}-e^{-\delta t_1})p_{x,r}(t_1,t_2)dt_2 + \int_{t_2\in S(t_1)}\frac{1}{2}[re^{-\delta t_2}-e^{-\delta t_1}]p_{x,r}(t_1,t_2)dt_2\Bigg]dt_1 \tag{A16}$$

where h denotes the joint probability density of x and rx. As in the case of delivery times, the identity

$$B_x(x,r)\equiv B_{rx}(rx,1/r) \tag{A17}$$

must hold for all x and r. This implies that (A16) will also hold in the case $r<1$. It is then easy to show that $rx\in\Sigma(x)\Leftrightarrow x\in\Sigma(rx)$, i.e. similarity relations in the prize dimension are symmetric.

A3. Proof of Lemma 1

Assume then that x_1 and x_2 are independent of t and Δ (so that $p_{t,\Delta}(x_1,x_2)=h(x_1,x_2)$ for all t and Δ). Then, from (A11), $B_t(t,\Delta)$ can be decomposed into

$$B_t(t,\Delta)=g(t,t+\Delta)\phi(\Delta) \tag{A18}$$

where ϕ is continuous for all $\Delta > 0$ and does not depend on t. We now show that ϕ is increasing for $\Delta > 0$. Take any pair Δ' and Δ'' with $\Delta' > \Delta'' > 0$. Then,

$$
\begin{aligned}
\phi(\Delta') - \phi(\Delta'') &= \\
\int_{x_1} \left\{ \int_{x_2 \in \Sigma_c(x_1) \cap [x_1, x_1 e^{\delta\Delta'}]} (x_1 - e^{-\delta\Delta'} x_2) h(x_1, x_2) dx_2 \right. &- \\
\int_{x_2 \in \Sigma_c(x_1) \cap [x_1, x_1 e^{\delta\Delta''}]} (x_1 - e^{-\delta\Delta''} x_2) h(x_1, x_2) dx_2 &+ \\
\left. \frac{1}{2}(e^{-\delta\Delta''} - e^{-\delta\Delta'}) \int_{x_2 \in \Sigma(x_1)} x_2 h(x_1, x_2) dx_2 \right\} dx_1 &= \\
\int_{x_1} \left\{ \int_{x_2 \in \Sigma_c(x_1) \cap [x_1 e^{\delta\Delta''},\ x_1 e^{\delta\Delta'}]} (x_1 - e^{-\delta\Delta'} x_2) h(x_1, x_2) dx_2 \right. &+ \\
\left. (e^{-\delta\Delta''} - e^{-\delta\Delta'}) \left[\int_{x_2 \in \Sigma_c(x_1) \cap [x_1, x_1 e^{\delta\Delta''}]} x_2 h(x_1, x_2) dx_2 + \frac{1}{2} \int_{x_2 \in \Sigma(x_1)} x_2 h(x_1, x_2) dx_2 \right] \right\} dx_1 &\geq \\
\int_{x_1} \left\{ \int_{x_2 \in \Sigma_c(x_1) \cap [x_1 e^{\delta\Delta''},\ x_1 e^{\delta\Delta'}]} \left(x_1 - e^{-\delta\Delta'} \left(x_1 e^{\delta\Delta'} \right) \right) h(x_1, x_2) dx_2 \right. &+ \\
\left. (e^{-\delta\Delta''} - e^{-\delta\Delta'}) \left[\int_{x_2 \in \Sigma_c(x_1) \cap [x_1, x_1 e^{\delta\Delta''}]} x_2 h(x_1, x_2) dx_2 + \frac{1}{2} \int_{x_2 \in \Sigma(x_1)} x_2 h(x_1, x_2) dx_2 \right] \right\} dx_1 &= \\
(e^{-\delta\Delta''} - e^{-\delta\Delta'}) \int_{x_1} \left[\int_{x_2 \in \Sigma_c(x_1) \cap [x_1, x_1 e^{\delta\Delta''}]} x_2 h(x_1, x_2) dx_2 + \frac{1}{2} \int_{x_2 \in \Sigma(x_1)} x_2 h(x_1, x_2) dx_2 \right] dx_1 &> 0
\end{aligned}
\tag{A19}
$$

Notice that the strictness of the last inequality comes from the fact that $\Sigma(x_1)$ and $\Sigma_c(x_1) \cap [x_1, x_1/e^{\delta\Delta''}]$ cannot be both empty. This shows that $\phi(\Delta)$ is strictly increasing for all $\Delta > 0$. Consider now what happens if Δ is negative. Notice first that the symmetry argument used for B_t in (A12) can be adapted to show that $\phi(-\Delta) \equiv e^{\delta\Delta}\phi(\Delta)$ for all $\Delta > 0$. It then follows that, for $\Delta > 0$, $\phi(-\Delta)$ must be continuous and increasing in Δ, which in turn implies that ϕ is continuous and decreasing for all negative Δ.

Finally, consider $\phi(\Delta)$ at $\Delta = 0$. Since, as already mentioned, $B_t(t, \Delta) = 0$ for all t and g has full support, it must be that $\phi(0) = 0$. Consider now $\lim_{\Delta \to 0^+} \phi(\Delta)$. The first term in (A11) converges to zero. The second term converges to

$$\frac{1}{2}\int_{x_1}\left[\int_{x_2\in\Sigma(x_1)}(x_1-x_2)h(x_1,x_2)dx_2\right]dx_1 \quad =$$

$$\text{(A20)}\quad \frac{1}{2}\left[\int_{x_1}\int_{x_2\in\Sigma(x_1)}x_1h(x_1,x_2)dx_2dx_1-\int_{x_1}\int_{x_2\in\Sigma(x_1)}x_2h(x_1,x_2)dx_2dx_1\right].$$

The above expression is zero whenever x_1 and x_2 are exchangeable. In order to show this, let $I(x_2;x_1)$, $I:(0,\overline{x})^2\to\{0,1\}$, be an indicator function taking value 1 if $x_2\in\Sigma(x_1)$ and 0 otherwise. Then, the first term in brackets in the second line of (A20) can be written as

$$\begin{aligned}\int_{x_2}\int_{x_1}x_1I(x_2;x_1)h(x_1,x_2)dx_1dx_2 &= \int_{x_2}\int_{x_1\in\Sigma(x_2)}x_1h(x_1,x_2)dx_1dx_2 = \\ \text{(A21)}\qquad \int_{x_1}\int_{x_2\in\Sigma(x_1)}x_2h(x_2,x_1)dx_2dx_1 &= \int_{x_1}\int_{x_2\in\Sigma(x_1)}x_2h(x_1,x_2)dx_2dx_1\end{aligned}$$

where the first equality follows from the fact that $x_2\in\Sigma(x_1)\Leftrightarrow x_1\in\Sigma(x_2)$, the second equality comes from relabeling x_1 as x_2 and x_2 as x_1, and the last equality follows from the fact that x_1 and x_2 are exchangeable in h. In turn, this implies that $\lim_{\Delta\to0^+}\phi(\Delta)=0$. One can then use the identity $\phi(-\Delta)\equiv e^{\delta\Delta}\phi(\Delta)$ to prove that the left limit is also zero. Since ϕ is continuous for all $\Delta\neq0$, this implies that ϕ is continuous everywhere.

Consider now the same for prizes. Assume that t_1 and t_2 are independent of x and r (so that $p_{x,r}(t_1,t_2)=g(t_1,t_2)$ for all x and r). Then, from (A16), B_x/x can be decomposed into

$$\text{(A22)}\qquad \frac{B_x(x,r)}{x}=h(x,rx)\mu(r)$$

We now show that μ is increasing for $r>1$. Take any pair r' and r'' with $r'>r''>1$.

Then,

$$
\begin{aligned}
&\mu(r') - \mu(r'') = \\
&\int_{t_1}\left[\int_{t_2\in S_c(t_1)\cap[t_1,t_1+\ln r'/\delta]}(r'^{-\delta t_2} - e^{-\delta t_1})g(t_1,t_2)dt_2 - \right.\\
&\left.\int_{t_2\in \S_c(t_1)\cap[t_1,t_1+\ln r''/\delta]}(r''^{-\delta t_2} - e^{-\delta t_1})g(t_1,t_2)dt_2 + \frac{1}{2}(r'-r'')\int_{t_2\in S(t_1)} e^{-\delta t_2}g(t_1,t_2)dt_2\right]dt_1 = \\
&\int_{t_1}\left[\int_{t_2\in S_c(t_1)\cap[t_1+\ln r'/\delta,t_1+\ln r''/\delta]}(r'^{-\delta t_2} - e^{-\delta t_1})g(t_1,t_2)dt_2 + \right.\\
&\left.(r'-r'')\int_{t_2\in S_c(t_1)\cap[t_1,t_1+\ln r'/\delta]} e^{-\delta t_2}g(t_1,t_2)dt_2 + \frac{1}{2}(r'-r'')\int_{t_2\in S(t_1)} e^{-\delta t_2}g(t_1,t_2)dt_2\right]dt_1 \geq \\
&\int_{t_1}\left[\int_{t_2\in S_c(t_1)\cap[t_1+\ln r'/\delta,t_1+\ln r''/\delta]}\left(r'\left(e^{-\delta t_1 - \ln r'}\right) - e^{-\delta t_1}\right)g(t_1,t_2)dt_2 + \right.\\
&\left.(r'-r'')\int_{t_2\in S_c(t_1)\cap[t_1,t_1+\ln r'/\delta]} e^{-\delta t_2}g(t_1,t_2)dt_2 + \frac{1}{2}(r'-r'')\int_{t_2\in S(t_1)} e^{-\delta t_2}g(t_1,t_2)dt_2\right]dt_1 = \\
&(r'-r'')\int_{t_1}\left[\int_{t_2\in S_c(t_1)\cap[t_1,t_1+\ln r'/\delta]} e^{-\delta t_2}g(t_1,t_2)dt_2 + \frac{1}{2}\int_{t_2\in S(t_1)} e^{-\delta t_2}g(t_1,t_2)dt_2\right]dt_1 > 0
\end{aligned}
\tag{A23}
$$

where, again, $S_c(t_1) \cap [t_1, t_1 + \ln r/\delta]$ and $S(t_1)$ cannot be both empty. This shows that μ is increasing for $r > 1$. As in the case of delivery times, we can use identity (A17) to obtain $\mu(r) \equiv \mu(1/r)$ for all $r > 1$. Since, for $r > 1$, $\mu(1/r)$ is increasing in r, it must be that μ is a decreasing function when r is lower than one. To conclude the proof, an analogous argument to the one presented above for ϕ can be made to show that $\mu(1) = 0$ and that μ is continuous if t_1 and t_2 are exchangeable.

A4. Proof of Proposition 2

Let us separate the case where the decision maker does not distinguish between x_1 and x_2 from the case where he perceives the difference. [We use the standard notation $\succ$, $\succeq$, and $\sim$ to denote strict preference, weak preference and indifference, respectively.]

a) Assume first $x_1 \approx x_2$. (i) If $t_1 \approx t_2$, then, since the length of similarity intervals increases with time (Proposition 1), we must have $t_1 + s \approx t_2 + s$. The decision maker's preferences are thus $(x_1, t_1) \sim (x_2, t_2)$ and $(x_1, t_1 + s) \sim (x_2, t_2 + s)$. (ii) If $t_1 \not\approx t_2$, then we may have $t_1 + s \not\approx t_2 + s$ or $t_1 + s \approx t_2 + s$. In the first case, $(x_1, t_1) \succ (x_2, t_2)$ and $(x_1, t_1+s) \succ (x_2, t_2+s)$. In the second case, $(x_1, t_1) \succ (x_2, t_2)$ and $(x_1, t_1+s) \sim (x_2, t_2+s)$. As a result, there is no preference reversal.

b) Hereafter, assume $x_1 \not\approx x_2$. (i) Again, since the length of the similarity intervals becomes larger as time increases, $t_1 \approx t_2 \Rightarrow t_1 + s \approx t_2 + s$. As a result, if $t_1 \approx t_2$, then $(x_1, t_1) \succ (x_2, t_2)$ and $(x_1, t_1 + s) \succ (x_2, t_2 + s)$. In this case, there is no preference reversal. (ii) If $t_1 \not\approx t_2$, then we may have $t_1 + s \not\approx t_2 + s$ or $t_1 + s \approx t_2 + s$. If $t_1 + s \not\approx t_2 + s$, the decision maker's preferences single out the fitness maximizing bundle throughout: $(x_1, t_1) \succ (x_2, t_2)$ and $(x_1, t_1+s) \succ (x_2, t_2+s)$ if $x_1 e^{-\delta t_1} > x_2 e^{-\delta t_2}$. Otherwise, if $x_1 e^{-\delta t_1} \leq x_2 e^{-\delta t_2}$, $(x_2, t_2) \succeq (x_1, t_1)$ and $(x_2, t_2 + s) \succeq (x_1, t_1 + s)$. In both cases there is no preference reversal. Suppose now that $t_1 + s \approx t_2 + s$. Then, the decision maker's preferences select the fitness maximizing bundle when comparing (x_1, t_1) to (x_2, t_2), but pick instead the bundle with the largest prize when comparing $(x_1, t_1 + s)$ to $(x_2, t_2 + s)$. In other words, $(x_1, t_1) \succ (x_2, t_2)$ if $x_1 e^{-\delta t_1} > x_2 e^{-\delta t_2}$ and $(x_2, t_2) \succeq (x_1, t_1)$ if $x_1 e^{-\delta t_1} \leq x_2 e^{-\delta t_2}$, while $(x_2, t_2 + s) \succ (x_1, t_1 + s)$. As a result, if $x_1 e^{-\delta t_1} > x_2 e^{-\delta t_2}$, preference reversal occurs. This shows that: a) for preference reversal to occur it is necessary that the earlier bundle generates higher fitness, b) preference reversal always takes the form of preferring the sooner/smaller bundle in the short term and preferring the later/larger bundle in the long term. $\square$

A5. Proof of Proposition 3

We distinguish the case where the decision maker perceives t_1 and t_2 as similar from the case where he perceives them as different.

a) Assume first $t_1 \approx t_2$. (i) If $x_1 \not\approx x_2$, then we must also have $\alpha x_1 \not\approx \alpha x_2$ given $\alpha > 1$. This follows from the fact that the size of the similarity intervals for the ratio between the two prizes increases as x decreases (Proposition 1). The decision maker's preferences in this case are thus $(x_1, t_1) \prec (x_2, t_2)$ and $(\alpha x_1, t_1) \prec (\alpha x_2, t_2)$, so that no preference reversal occurs. (ii) If $x_1 \approx x_2$, then we may have $\alpha x_1 \approx \alpha x_2$ or $\alpha x_1 \not\approx \alpha x_2$. In the first case the decision maker's preferences are: $(x_1, t_1) \sim (x_2, t_2)$ and $(\alpha x_1, t_1) \sim (\alpha x_2, t_2)$. In the second case they are: $(x_1, t_1) \sim (x_2, t_2)$ and $(\alpha x_1, t_1) \prec (\alpha x_2, t_2)$. In both cases, there is no preference reversal.

b) Hereafter, assume $t_1 \not\approx t_2$. (i) If $x_1 \not\approx x_2$, then again from Proposition 1 we must also have $\alpha x_1 \not\approx \alpha x_2$. In this case, the decision maker's preferences single out the fitness maximizing bundle throughout; In other words, we either have: $(x_1, t_1) \succ (x_2, t_2)$ and $(\alpha x_1, t_1) \succ (\alpha x_2, t_2)$ (if $x_1 e^{-\delta t_1} > x_2 e^{-\delta t_2}$) or we have $(x_1, t_1) \preceq (x_2, t_2)$ and $(\alpha x_1, t_1) \preceq (\alpha x_2, t_2)$ (if $x_1 e^{-\delta t_1} \leq x_2 e^{-\delta t_2}$), so that no reversal occurs. (ii) If $x_1 \approx x_2$, then we may have either $\alpha x_1 \approx \alpha x_2$ or $\alpha x_1 \not\approx \alpha x_2$. In the first case the decision maker's preferences are $(x_1, t_1) \succ (x_2, t_2)$ and $(\alpha x_1, t_1) \succ (\alpha x_2, t_2)$, so that no reversal occurs. In the second case, the decision maker's preferences favor the bundle with the earlier arrival time when comparing (x_1, t_1) to (x_2, t_2), and single out the fitness maximizing bundle when comparing $(\alpha x_1, t_1)$ to $(\alpha x_2, t_2)$; In other words, we either have: $(x_1, t_1) \succ (x_2, t_2)$ and $(\alpha x_1, t_1) \succeq (\alpha x_2, t_2)$ (if $x_1 e^{-\delta t_1} \geq x_2 e^{-\delta t_2}$) or we have $(x_1, t_1) \succ (x_2, t_2)$ and $(\alpha x_1, t_1) \prec (\alpha x_2, t_2)$ (if $x_1 e^{-\delta t_1} < x_2 e^{-\delta t_2}$). Hence, in the last case we have a reversal of preferences and this takes the form stated in the proposition (i.e. preferring the sooner/smaller bundle when choosing between small prizes and choosing the later/larger bundle when choosing between larger prizes). Moreover, as argued, the necessary condition for preference reversal is that the bundle with the greater prize generates higher fitness. □

A6. Proof of Proposition 4

We start by noticing two intermediate results that follow immediately from Proposition 1. First, notice that, for $k > 1$, if $t + \Delta$ is not in t's similarity interval, then $t + k\Delta$ is also outside t's similarity interval. Similarly, if $rx \not\approx x$, then $r^k x \not\approx x$.

In order to prove that $t \not\approx t + \Delta$ must hold for subadditivity, it is enough to show that $(x, t) \succ (rx, t + \Delta) \Rightarrow t \not\approx t + \Delta$. To see this, suppose by contradiction that $t \approx t + \Delta$. Then, either $x \approx rx$, in which case $(x, t) \sim (rx, t + \Delta)$, or $x \not\approx rx$, in which case $(x, t) \prec (rx, t + \Delta)$.

Similarly, in order to prove that $t \approx t + \Delta$ must hold for superadditivity, suppose by contradiction that $t \not\approx t + \Delta$. We then have two possibilities. (1) $x \approx rx$, in which case $(x,t) \succ (rx, t+\Delta)$, or (2) $x \not\approx rx$, in which case the direction of preferences between (x,t) and $(rx, t+\Delta)$ depends on which bundle has the highest fitness. Note however that, as already established, $t \not\approx t+\Delta \Rightarrow t \not\approx t+k\Delta$ and $x \not\approx rx \Rightarrow x \not\approx r^k x$. The direction of preferences between (x,t) and $\left(r^k x, t+k\Delta\right)$ thus singles out the fitness maximizing bundle in this case too. As a result, we cannot have a preference reversal. This proves that a necessary condition for superadditivity is $t \approx t+\Delta$.

We now derive a parameter configuration such that the model generates subadditivity for given prizes and delivery times. This involves 4 steps

1) Pick values for r and Δ satisfying

(A24)
$$r^k e^{-\delta k \Delta} > 1$$

so that the bundle $(xr^k, t+k\Delta)$ generates higher fitness than (x,t).

2) Given Δ, pick (suitably small) values for t and c such that

(A25)
$$c < e^{-\delta t}\frac{\phi(\Delta)}{\overline{t}^2}$$

so that $t \not\approx t+\Delta$.

3) Given c and r, pick a value x_ϵ for x such that, for $\epsilon > 0$ arbitrarily small,

(A26)
$$\frac{x_\epsilon}{\overline{x}^2}\mu(r) + \epsilon = c$$

so that $rx_\epsilon \approx x_\epsilon$.

4) Notice now that, given any $k > 1$, there exists $\epsilon > 0$ sufficiently small that

(A27)
$$\frac{x_\epsilon}{\overline{x}^2}\mu(r^k) > c$$

which implies that $r^k x_\epsilon \not\approx x_\epsilon$.

It is then clear that, since $rx_\epsilon \approx x_\epsilon$, the decision maker will prefer the bundle with the shortest waiting time when choosing between (x_ϵ, t) and $(rx_\epsilon, t+\Delta)$, so that $(x_\epsilon, t) \succ (rx_\epsilon, t+\Delta)$. However, since $r^k x_\epsilon \not\approx x_\epsilon$, he will prefer the fitness maximizing bundle when choosing between (x_ϵ, t) and $(r^k x_\epsilon, t+k\Delta)$, so that $(x_\epsilon, t) \prec (r^k x_\epsilon, t+k\Delta)$. As a result, preferences display subadditivity.

The steps required to generate superadditivity are symmetric to those shown above for subadditivity. First, we pick r and Δ such that $r^k e^{-\delta k\Delta} < 1$ so that $(xr^k, t+k\Delta)$ generates lower fitness than (x, t). Then, we can choose x and c such that $x\mu(r)/\overline{x}^2 > c$, so that $rx \not\approx x$. Next, for $\epsilon > 0$ small, we choose $t = t_\epsilon$ such that

$$e^{-\delta t_\epsilon} \frac{\phi(\Delta)}{\overline{t}^2} + \epsilon = c \tag{A28}$$

so that $t_\epsilon \approx t_\epsilon + \Delta$. Given a value of k, one can always find ϵ small enough such that $t_\epsilon \not\approx t_\epsilon + k\Delta$. As a result, when faced with (x, t_ϵ) and $(rx, t_\epsilon + \Delta)$, the decision maker will prefer the second bundle since t_ϵ and $t_\epsilon + \Delta$ are perceived as similar. However, when faced with (x, t_ϵ) and $(r^k x, t_\epsilon + k\Delta)$, he will prefer the first bundle since t_ϵ and $t_\epsilon + k\Delta$ are perceived as different. As a result, preferences display superadditivity.

We now prove the last statement in the proposition, namely that for both sub- and superadditivity, the decision maker's preferences when facing the long interval treatment single out the fitness maximizing bundle. Suppose by contradiction that this is not the case. Then, necessarily, either $r^k x \approx x$ or $t + k\Delta \approx t$, or both. If $r^k x \approx x$, then $rx \approx x$. As a result, the decision maker would consistently prefer the early bundle in both cases, so that neither super nor subadditivity is possible. If $t + k\Delta \approx t$, then $t + \Delta \approx x$. Hence, the decision maker would consistently prefer the larger bundle in both cases, so that no super/subadditivity is possible. Finally, if both $r^k x \approx x$ and $t + k\Delta \approx t$, then the decision maker would be consistently indifferent in both cases. $\square$

Supporting material for informal claims made in the main body (for online publication)

B1. Benhabib et al. (2009)

In this section, we argue that our model is consistent with empirical evidence on the magnitude effect obtained by eliciting indifference points. Consider the question (asked e.g. in Behabib et al. 2009) "What amount of money y would make you indifferent between x today and y in τ days?" where x is equal to 10/20/30 etc. dollars and τ is equal to 3 days, 1 week, 2 weeks, 1 month etc. depending on the treatment. It is conceivable that nearly all subjects are able to perceive a difference between today and 3 days from now (and any other date further away in the future). Hence, we assume at the outset that $\tau \not\approx 0$. Note that, in terms of fitness, the amount y that makes (y, τ) equivalent to $(x, 0)$ is equal to $xe^{\delta\tau}$. We thus have two possibilities; (a) If $e^{\delta\tau} \geq r^+(x)$, so that $e^{\delta\tau}$ lies outside x's similarity interval, then the individual will select $y = xe^{\delta\tau}$; (b) If $e^{\delta\tau} < r^+(x)$, the individual cannot distinguish between $xe^{\delta\tau}$and x. Any bundle (y, τ) where $y/x \leq r^+(x)$ is seen as indistinguishable to a bundle proposing x in τ days and is thus considered strictly worse than $(x, 0)$. However, the decision maker also considers any bundle (y, τ), where $y/x > r^+(x)$, as strictly preferable to $(x, 0)$. In other words, there is no value y that makes the individual truly indifferent between (y, τ) and $(x, 0)$. We conjecture that, in this case, the individual will select $y = xr^+(x)$, namely the smallest prize at τ that he prefers to x now. This is the value of the prize where his preferences over the two bundles change direction.

From Proposition 1, we know that $r^+(x)$ is decreasing in x. This implies that, keeping everything else equal, when x is large it is more likely that the individual selects $y = xe^{\delta\tau}$, while when x is small it is more likely that the individual selects $xr^+(x)$. Consider now a researcher who assumes that $u(x, 0) = x$ and $u(y, \tau) = ye^{-d(y,\tau)\tau}$ where $d(y, \tau)$ is the unknown discount rate to be estimated. When x is large, the researcher will find $e^{-d(y,\tau)\tau} = \frac{x}{xe^{\delta\tau}}$, i.e. $d(y, \tau) = \delta$. When x is small, the researcher will find $e^{-d(y,\tau)\tau} = x/r^+(x)x = 1/r^+(x) < e^{-\delta\tau}$, so that $d(y, \tau) > \delta$. The researcher will thus conclude that the discount rate decreases with the amount to be discounted.

B2. Read (2001), Scholten and Read (2006)

Consider the following class of experiments (Read 2001, Scholten and Read 2006, Dohmen et al. 2012). People are asked to compare (x, t) (smaller/sooner) with $(y, t + \Delta)$ (larger/later); The value of y starts low and is progressively increased. The researchers identify the switching point where the agent moves from preferring the smaller/sooner to the larger/later bundle, and elicit the implied discount factor. The results are then compared with those of other treatments with a different interval length, $k\Delta$, for some $k > 1$. From a fitness perspective, the value of y that makes (x, t) equivalent to $(y, t + \Delta)$ is $xe^{\delta\Delta}$.

(i) Suppose $t \not\approx t+\Delta$ (which, as usual, implies $t \not\approx t+k\Delta$). Generally, given any interval Δ, we have two possibilities. Either $e^{\delta\Delta} \geq r^+(x)$, so that $xe^{\delta\Delta}$ is not in x's similarity interval, or $e^{\delta\Delta} < r^+(x)$, so that $e^{\delta\Delta}$ is in x's similarity interval. In the first case, the agent will switch at $y = xe^{\delta\Delta}$ (or close to that amount if that option is not available). In the second case the agent will strictly prefer $(y, t + \Delta)$ to (x, t) for $y \geq xr^+(x)$, while for $y < xr^+(x)$ he will strictly prefer (x, t) to $(y, t + \Delta)$. The switching point is thus $y = xr^+(x)$. The same principle applies to the larger interval treatments. When $t \not\approx t+\Delta$, we thus have three possibilities

1) $e^{\delta k\Delta} > e^{\delta\Delta} \geq r^+(x)$, so that subjects select $y = xe^{\delta\Delta}$ in the short interval treatment and $y = xe^{\delta k\Delta}$ in the long interval treatment;

2) $r^+(x) \geq e^{\delta k\Delta} > e^{\delta\Delta}$, so that subjects select $y = xr^+(x)$ in both treatments;

3) $e^{\delta k\Delta} \geq r^+(x) > e^{\delta\Delta}$, so that subjects select $y = xr^+(x)$ in the short interval treatment and $y = xe^{\delta k\Delta}$ in the long interval treatment.

Only cases (2) and (3) may generate interval length effects. Take first case (2), and consider a researcher who assumes that $u(y, \tau) = ye^{-d\tau}$ where d is the (interval length dependent) unknown discount rate to be estimated. In the short interval treatment, the observed value d_{si} solves $e^{d_{si}\Delta} = r^+(x)$, while in the long interval treatment it solves $e^{d_{li}k\Delta} = r^+(x)$. Since $r^+(x) > 1$ and $k > 1$,

$$d_{li} = \frac{\ln r^+(x)}{k\Delta} < \frac{\ln r^+(x)}{\Delta} = d_{si} \tag{B1}$$

The researcher thus concludes that a larger discount factor (a smaller rate) is used for the long interval relative to the short interval (subadditivity). Consider now case (3). In the long interval treatment, $d_{li} = \delta$. In the short interval treatment, d_{si} solves $e^{d_{si}\Delta} = r^+(x) > e^{\delta\Delta}$, i.e. $d_{si} > \delta$. Again, the researcher obtains $d_{li} < d_{si}$, i.e. subadditivity.

(ii) Suppose now that $t \not\approx t + k\Delta$ but $t \approx t + \Delta$, so that the decision maker perceives the difference in delivery times for the long interval but not for the short interval. We maintain $r^+(x) < e^{\delta k\Delta}$, so that the decision maker will choose $y = xe^{\delta k\Delta}$ when facing the long interval. This implies $d_{li} = \delta$. Consider now the short interval. Since the decision maker does not distinguish between t and $t + \Delta$, he may in principle pick any y in x's similarity interval, i.e. any y on offer that does not exceed $xr^+(x)$. If he chooses $y < xe^{\delta\Delta}$, we will have $d_{si} < \delta = d_{li}$, i.e. superadditivity. If $e^{\delta\Delta}$ is not much smaller than $r^+(x)$, this may hold for a majority of subjects. As a result, the researcher may conclude that most subjects display superadditivity. Note that, consistent with Scholten and Read's (2006) findings, this also implies that, if Δ is progressively increased, the researcher should first obtain superadditivity and then subadditivity.

B3. The (σ, δ) model (Manzini and Mariotti, 2006)

We now prove two claims made in the text. First, that only the outcome-prominence version of the (σ, δ) model delivers time inconsistency in the form that has been empirically documented. Second, that, in the outcome-prominence version, interval length effects only take the form of superadditivity. Consider two comparisons, (i) between (x_1, t_1) and (x_2, t_2) and (ii) between $(x_1, t_1 + s)$ and $(x_2, t_2 + s)$, for some $s > 0$. For notational consistency, we maintain our convention of denoting with δ the instantaneous discount rate (so that $e^{-\delta}$ is the discount factor). The difference in utilities is then

$e^{-\delta t_1}\left|e^{-\delta(t_2-t_1)}x_2 - x_1\right|$ in case (i), and $e^{-\delta(t_1+s)}\left|e^{-\delta(t_2-t_1)}x_2 - x_1\right|$ in case (ii). We thus have 3 possibilities. (1) $e^{-\delta t_1}\left|e^{-\delta(t_2-t_1)}x_2 - x_1\right| < \sigma$, (2) $e^{-\delta(t_1+s)}\left|e^{-\delta(t_2-t_1)}x_2 - x_1\right| > \sigma$ and (3) $e^{-\delta t_1}\left|e^{-\delta(t_2-t_1)}x_2 - x_1\right| > \sigma > e^{-\delta(t_1+s)}\left|e^{-\delta(t_2-t_1)}x_2 - x_1\right|$. Time preference reversal may only occur in case (3). There, the agent prefers the utility maximizing bundle when comparing (x_1, t_1) and (x_2, t_2), and uses heuristics when comparing $(x_1, t_1 + s)$ and $(x_2, t_2 + s)$. Note that, for time preference reversal to occur, it is necessary that, when comparing $(x_1, t_1 + s)$ and $(x_2, t_2 + s)$, the agent prefers the later/larger bundle. This may only occur in the outcome-prominence case.

Consider now interval length effects and suppose, similar to Section III, that the decision maker faces two choices. The first is between (x, t) and $(rx, t + \Delta)$ and the second is between (x, t) and $(r^k x, t + k\Delta)$. Suppose first that $1 > re^{-\delta\Delta}$, i.e. the sooner/smaller bundle is utility maximizing in both cases. Note that, since $re^{-\delta\Delta} < 1$, $1 - re^{-\delta\Delta} < 1 - r^k e^{-\delta k\Delta}$ follows. We have three possibilities. (i) $1 - re^{-\delta\Delta} < 1 - r^k e^{-\delta k\Delta} < \sigma$; (ii) $\sigma < 1 - re^{-\delta\Delta} < 1 - r^k e^{-\delta k\Delta}$ and (iii) $1 - re^{-\delta\Delta} < \sigma < 1 - r^k e^{-\delta k\Delta}$. Clearly enough, there cannot be any preference reversal under case ii). Under case i), the decision maker chooses the earlier bundle in both cases (if delivery time is prominent) or the later bundle in both cases (if outcome is prominent). As a result, no preference reversal arises also in this case. The only case in which we may obtain preference reversal is (iii). In that case, when comparing (x, t) and $(rx, t + \Delta)$ the agent uses the outcome-based heuristics and prefers the later/larger bundle. When comparing (x, t) and $\left(r^k x, t + k\Delta\right)$ the agent prefers the utility maximizing bundle, i.e. the sooner/smaller one. Hence, $(x, t) \prec (r, t + \Delta)$, and $(x, t) \succ \left(r^k x, t + k\Delta\right)$, i.e., we have superadditivity.

Suppose now that $1 \leq re^{-\delta\Delta}$, i.e. the later/larger bundle is utility maximizing in both cases. Note that, since $re^{-\delta\Delta} \geq 1$, $1 - re^{-\delta\Delta} \geq 1 - r^k e^{-\delta k\Delta}$ follows. We have three possibilities. (i) $1 - re^{-\delta\Delta} \geq 1 - r^k e^{-\delta k\Delta} > \sigma$; (ii) $\sigma > 1 - re^{-\delta\Delta} \geq 1 - r^k e^{-\delta k\Delta}$ and (iii) $1 - re^{-\delta\Delta} > \sigma > 1 - r^k e^{-\delta k\Delta}$. Again, the only candidate preference reversal is (iii). However, even in that case, preference reversal does not occur. When comparing (x, t) and $(rx, t + \Delta)$ the agent prefers the utility maximizing bundle, i.e. the later/larger one. When comparing (x, t) and $\left(r^k x, t + k\Delta\right)$ the agent uses the outcome-based heuristic and therefore again prefers the later/larger bundle. This proves that preference reversal may only take the form of superadditivity.

B4. The fixed cost model (Benhabib et al. 2009)

Here we show that the fixed cost model delivers subadditivity but not superadditivity. Suppose, again, that the decision maker faces two choices. The first is between (x, t) and $(rx, t+\Delta)$ and the second is between (x, t) and $(r^k x, t+k\Delta)$. Consider first superadditivity, i.e.: (1) $x < rx\rho^{\Delta} - b$ and (2) $x > r^k x\rho^{k\Delta} - b$. Note that (1) requires $b < rx\rho^{\Delta} - x$. By contrast, (2) requires $b > r^k x\rho^{k\Delta} - x$. Consistency between (1) and (2) requires $rx\rho^{\Delta} > r^k x\rho^{k\Delta}$, i.e. $r\rho^{\Delta} < 1$. This however contradicts (1).

Consider now subadditivity, i.e.: (3) $x > rx\rho^{\Delta} - b$ and (4) $x < r^k x\rho^{k\Delta} - b$. Clearly enough, consistency between (3) and (4) requires $r^k x\rho^{k\Delta} > rx\rho^{\Delta}$, i.e. $r\rho^{\Delta} > 1$. There is no contradiction.