ELSEVIER

Progress in Neurobiology 84 (2008) 284–315

progress in Neurobiology

www.elsevier.com/locate/pneurobio

Is a bird in the hand worth two in the future? The neuroeconomics of intertemporal decision-making

Tobias Kalenscher [*], Cyriel M.A. Pennartz

Animal Physiology and Cognitive Neuroscience, Swammerdam Institute for Life Sciences, Faculty of Science, University of Amsterdam, Kruislaan 320, 1098 SM Amsterdam, The Netherlands

Received 3 May 2007; received in revised form 28 November 2007; accepted 29 November 2007

Abstract

When making intertemporal decisions, i.e., decisions between outcomes occurring at different instants in time, humans and animals prefer rewards with short-term availability over rewards that become available in the long run. Discounted utility theory (DUT) is an influential normative model for intertemporal decisions that attempts to capture preference over time. It prescribes which is the best decision to take with respect to consistent, coherent and optimal choice. Over the last few decades, DUT's descriptive validity has been critically challenged. Empirical studies found systematic violations of several of DUT's assumptions, including time-consistency of preferences, stationarity, constant discounting and utility maximisation. To account for these anomalies, alternative models have been devised in various academic disciplines, including economics, psychology, biology, philosophy, and most lately, cognitive neuroscience. This article reviews the most recent literature on the behavioural and neural processes underlying intertemporal choices, and elucidates to which extent these findings can be used to explain violations of DUT's assumptions. In the first three sections, DUT is introduced, and behavioural anomalies are discussed. The fourth part focuses on the neuroscience of intertemporal choice, including its functional neuroanatomy, attempts to find a discounted value signal in the brain, and recent efforts to identify neural mechanisms producing time-inconsistencies. In the last section, the computational literature on neural learning mechanisms is reviewed. Then, a new, biologically plausible computational model of intertemporal choice is proposed that is able to explain many of the behavioural anomalies. The implications of these results help to understand why humans and animals frequently decide irrationally and to their long-term economic disadvantage, and which neural mechanisms underly such decisions.

Keywords: Neuroeconomics; Reward; Decision-making; Intertemporal decisions; Delay discounting; Risk; Impulsivity; Self-control; Dopamine; Glutamate; Temporal difference learning; Neural replay

Contents

Abbreviations: ACC, anterior cingulate cortex; BLA, basolateral amygdala; BOLD, blood-oxygen-level dependent; cAMP, cyclic adenosine monophosphate; CS, conditioned stimulus; DLPFC, dorsolateral prefrontal cortex; EUT, expected utility theory; HSAT, Hebbian synapses with adaptive thresholds model; LTD, long-term depression; LTP, long-term potentiation; mPFC, medial prefrontal cortex; NAc, nucleus accumbens; NCL, nidopallium caudolaterale; NMDA, *n*-methyl D-aspartate; OFC, orbitofrontal cortex; PFC, prefrontal cortex; PSTH, peri-stimulus time histogram; RPM, reinforcement processing module; STN, subthalamic nucleus; TD, temporal difference; VMPFC, ventromedial prefrontal cortex.

[*] Corresponding author. Tel.: +31 20 525 7637; fax: +31 20 525 7709.
E-mail address: T.Kalenscher@uva.nl (T. Kalenscher).

0301-0082/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pneurobio.2007.11.004

1. Time and probability

1.1. Decisions under risk and intertemporal decisions

Your door bell rings, and when you open it, an insurance salesman smiles into your face. He wants to sell you two different insurance certificates. The first certificate is a fire insurance. It would cost you only $10 per month, and would cover all the damage to your house in the unlikely, but shattering case it was destroyed by a terrible fire. The second policy is a pension fund. The premium is $100 each month, but you would receive a guaranteed life-long retirement pay of at least $700 per month once you reach the age of 65.

Those two examples nicely illustrate two problems in choice theory that keep decision researchers from many different scientific disciplines busy, including psychology, biology, neuroscience, economics, law, politics and philosophy: decision-making under risk and intertemporal decision-making. When you deliberate whether to buy or reject the fire policy offer, you will try to estimate the probability that your house could indeed be destroyed by a fire, and trade off this small risk with the devastating problem you would be facing if it actually happened. You are, hence, making a decision under risk, i.e., you are trying to decide between probabilistic outcomes.[1] The second type of decision-making, intertemporal choices, is illustrated by the pension fund example: when thinking about investing into the retirement provision, you will face little risk or uncertainty, as the insurance company is reliable and will pay out the pension when you reach the age. However, your decision would involve calculating and trading-off choice outcomes that would be realised at different points in time, i.e., you would have to invest money now to obtain benefits that are yet to come. On the other hand, if you decided against the retirement provision, you could save the monthly $100 premium. This would put you in a better financial position to treat yourself to things that you fancy now, for example an expensive dinner each month. Hence, your choice depends on whether you are willing to forgo short-term benefits in order to invest into an option that is probably more reasonable in the long run. Accordingly, intertemporal decision making can be defined as choices between outcomes that occur at different points in time.

This review will focus on the neurobiology underlying the second class of choices, intertemporal decisions, although the first class, decisions under risk, will be mentioned frequently throughout this article as well. For a more detailed overview about the psychology and neurobiology of decisions under risk, refer to the existing literature (for example, Von Neumann and Morgenstern, 1944; Friedman and Savage, 1948, 1952; Kahneman and Tversky, 1979; Platt and Glimcher, 1999; Montague and Berns, 2002; Fiorillo et al., 2003; Barraclough et al., 2004; Dorris and Glimcher, 2004; Glimcher, 2004; Glimcher and Rustichini, 2004; McClure et al., 2004; Schultz, 2004; Hsu et al., 2005; Knutson et al., 2005; McCoy and Platt, 2005; Tobler et al., 2005; Trepel et al., 2005; Sanfey et al., 2006).

[1] As always, this everyday example may be somewhat misleading since it may only inappropriately capture the concept of decisions under risk. Although decisions under risk obviously entail choices between probabilistic outcomes, many researchers actually use the term 'risk' strictly in the sense of outcome variance in a multi-choice situation, i.e., a certain outcome has a variance of 0, a risky option has an outcome with a variance >0 (e.g., Kacelnik, 1997; Kacelnik and Bateson, 1996, 1997).

1.2. Probabilistic and dated outcomes

The conflict in the introductory example about whether or not to invest into a pension fund entails the decision between dated outcomes. In general, a substantial body of evidence suggests that, provided the costs for all options are identical, the preference for an immediate or a temporally remote outcome is a function of the value of the respective outcomes and their delays, i.e., the time until they can be realised (McDiarmid and Rilling, 1965; Rachlin and Green, 1972; Ainslie, 1975; Mazur, 1984, 1987, 1988; Grossbard and Mazur, 1986; Logue, 1988; Benzion et al., 1989; Green et al., 1994, 1997; Evenden and Ryan, 1996; Evenden, 1999; Frederick et al., 2002; Reynolds et al., 2002; Kalenscher et al., 2005b, 2006a). For example a given reward, delivered after a long delay, is less attractive than the same reward delivered after a short delay. The process of systematically devaluating outcomes over time is called temporal discounting.

Why do we discount the future? Take the introductory scenario: obviously, you may decide against the pension fund because you do not want to wait several decades to receive the pay-out. It is equally possible, though, that you may reject the pension fund because you are uncertain whether you'll ever reach the pension age, or because you have doubts about the financial integrity of the insurance company, despite its good reputation. In this case, you would treat the intertemporal decision as a choice between uncertain, and not between delayed outcomes. Accordingly, many researchers implicitly or openly presumed that delay in intertemporal decision-making affects choice behaviour in a similar way as probability does in decision-making under risk (Mischel and Grusec, 1967; Kagel et al., 1986; Benzion et al., 1989; Prelec and Loewenstein, 1991; Rachlin et al., 1991; Keren and Roelofsma, 1995; Kacelnik and Bateson, 1996; Green and Myerson, 1996, 2004; Kacelnik, 1997; Sozou, 1998; Frederick et al., 2002; Yi et al., 2006). In other words, humans and other animals may equate temporal distance with uncertainty: a temporally proximal reward may be preferred over a temporally distant reward in the same way as a likely reward is preferred over an unlikely reward, since delayed benefits may be lost during waiting time, are less likely to be realised, and cannot be put to use until they are realised. Some authors have turned this logic around (Rachlin et al., 1986; Mazur, 1989; Hayden and Platt, 2007) and argued that a probabilistic reward in a multi-choice situation may be construed as a certain reward with a variable delay, since subjects will, almost with full certainty, receive the probabilistic reward eventually if they consistently stick with it: if not on the current trial, then on a future trial. Hence, probabilistic and delayed rewards may be cognitively treated in an identical fashion, and may therefore recruit similar neural mechanisms. However, despite some support for the theories of shared cognitive and neural mechanisms (Mobini et al., 2002; Hayden and Platt, 2007), behavioural results (Rachlin et al., 1986, 1991; Snyderman, 1987; Mazur, 1989) as well as evidence from lesion and psychopharmacological studies (Mobini et al., 2000; Cardinal and Howes, 2005; Acheson et al., 2006) are inconsistent and generally point more towards a dissociation of mechanisms.

2. Rational intertemporal decisions

2.1. (Ir)rational decisions

A clear-cut definition of rationality that is accepted across all academic disciplines is difficult to find (cf. Kacelnik, 2006). In economics, rationality can be defined as internal consistency with manifest preference ordering. Normative models of rational decision-making, i.e., models prescribing which is the best decision to take with respect to consistent, coherent and optimal choice, are concerned with how people's overt preferences can be transformed into internal value relations and vice versa. These models contain a list of axioms that must be met to allow this transformation (see Section 2.2 for details). An ideal, rational decision maker should obey these axioms. Hence, this definition of rationality implies conformity with a well-defined normative model. In this sense, any deviation from the normative ideal would be classified as 'irrational'. To understand irrational behaviour, it is therefore important to understand why people frequently exhibit preferences that deviate from this normative ideal. The rationality definition in economics is broadly in harmony with some philosophical theories that conceptualise 'rationality' in the sense of veracity or conformity with one's own internal moral, ethical, social and economic value system (Raz, 1999). Most rational choice models have in common that they presume a motivation for maximisation of some currency under given constraints, for example, the maximisation of energy intake as a proxy for Darwinian fitness (in biology), or maximisation of utility (in economics).[2] Throughout this review, we will use the term 'rational' in the sense of conformity to a normative model of choice.

2.2. Expected utility theory as a normative framework for decision-making under risk

Before moving on to normative models of intertemporal decisions, it is necessary to briefly outline one of the most influential normative frameworks of decisions under risk, called expected utility theory (EUT). Utility theories in general, both for the domain of time and risk, make basic assumptions (axioms) about the elements of the decision space and the preference relations of the decision maker with respect to these elements. From these axioms the theories deduce statements about how the preference relations, as observed from the actual choices, can be transformed into utility relations (numbers indicating the subjective values of the commodities). Only if an

[2] The term 'utility' is widely used in economics and can be understood as a measure of relative satisfaction or gratification. Normally, choice models in economics do not exclusively refer to the maximisation of monetary gains, but also refer to more abstract benefits, such as obtaining pleasure from engaging in a favourite recreational activity, or enjoying one's favourite food.

agent's preference relation satisfies the axioms, can a utility U be assigned to each alternative choice option.

EUT was axiomised more than 60 years ago. Its axioms include completeness, transitivity, continuity and independence of preferences (see Von Neumann and Morgenstern, 1944 for details). Simply speaking, completeness demands that there are preferences across alternatives, transitivity requires that preferences are ordered hierarchically, continuity requires that there exists some probability such that the decision-maker is indifferent between the most preferred and the least preferred outcome, and independence demands that preference orders do not reverse if a non-preferred alternative is added to the option set.

In essence, EUT posits that a decision maker chooses between risky prospects based on the utility of final asset positions. This means that the decision maker chooses the option yielding the highest expected utility of the final wealth state after integrating the options' prospects with the current wealth. Expected utility is measured in arbitrary units, and is the sum of the probability-weighted expected subjective values of all possible outcomes. The subjective values are not necessarily identical to their objective values, as current wealth, biases, individual preferences, contexts and environmental and social influences may affect their valuation (Friedman and Savage, 1948, 1952; Bernoulli, 1954; cf. Kahneman and Tversky, 1979; Powell, 2003; Glimcher and Rustichini, 2004). Utility functions are generally believed to be concave for gains, i.e., with increasing assets, the utility function increases sublinearly and in a decelerating fashion. This concavity is considered to underly risk aversion (Friedman and Savage, 1948; Bernoulli, 1954). For the sake of convenience, throughout this review, we will use the term 'utility', which is more common in economics, synonymously with the term 'subjective value', which is more common in psychology and biology, although we acknowledge that not all authors in economics, psychology and biology may agree on the equivalence of the two terms.

2.3. Discounted utility theory as a normative framework for intertemporal decisions

The discounted utility model (DUT) is the equivalent to EUT in the domain of time. It provides an axiomised normative framework to account for intertemporal decisions and parallels EUT in many regards. For example, both models assume that decision makers choose between options based on a weighted sum of utilities, with either probabilities (EUT) or temporal discount factors (DUT) as weights. Accordingly, the intertemporal utility function in DUT can be described by the following equation:

$$U_t(c_t, \ldots, c_T) = \sum_{k=0}^{T-t} D(k)u(c_{t+k}) \tag{1}$$

where $u(c_{t+k})$ is the instantaneous utility that a consumption (reward) c will have at timepoint $t + k$, T is the time horizon, and $D(k)$ is the individual discount function, i.e., the weight by which the instantaneous utility at the later timepoint $t + k$ is discounted from the perspective of timepoint t. As discussed below, DUT assumes that the discount function $D(k)$ is exponential (Samuelson, 1937). Importantly, DUT posits that intertemporal decision-making can be described by a single discount function $D(k)$.

DUT has been first proposed more than 70 years ago by Samuelson (1937), and has been axiomised and further developed throughout the 20th century (Koopmans, 1960; Lancaster, 1963; Fishburn and Rubinstein, 1982; cf. Prelec and Loewenstein, 1991; Loewenstein, 1992; Frederick et al., 2002). DUT demands that preferences between delayed outcomes satisfy the following axioms:

- monotonicity of time preference,
- completeness of time preference,
- intertemporal transitivity,
- continuity of time preference
- intertemporal independence,
- stationarity.

Furthermore, the theory assumes:

- constant discount rate,
- maximisation of utility rate.

Monotonicity in the domain of uncertainty means that stochastically dominating prospects are preferred over stochastically dominated prospects (a stochastically dominating prospect is a prospect that can be ranked as superior to the dominated prospect). Monotonicity of time preference (Lancaster, 1963) is the equivalent in the domain of time and holds that

$$A(t_1) \geq A(t_2), \quad \text{if, and only if,} \quad t_2 \geq t_1 \tag{2}$$

This means that commodity A, available at timepoint t_1, will be preferred over A, available at timepoint t_2, if, and only if $A(t_2)$ occurs later than $A(t_1)$. This axiom formalises the empirical finding that a short-term reward is preferred over a long-term reward.

Completeness, transitivity, and continuity are basically equivalent to the axioms in EUT, only translated to the domain of time. For example, intertemporal transitivity implies that, if commodity A, delivered at timepoint t_1, is preferred to commodity B, delivered at timepoint t_2, and B at time t_2 is preferred to C at time t_3, then A at time t_1 will be preferred to C at t_3. These axioms will not be further discussed here. As concerns consumption independence in intertemporal choice: this axiom states that preferences for consumptions (rewards) should not be affected by the nature of consumptions in periods in which consumption is identical. This means that the utility of a reward should be independent of whether a reward was already experienced in the past, or will be experienced at another time. Even Samuelson and Koopmans acknowledged that this assumption is of limited validity (cf. Frederick et al., 2002), because an agent's current preference between, say, pizza and steak is most certainly affected by whether he had pizza already for the last few days.

Stationarity posits that

$$\text{If}\quad (A,t) \sim (B,t+\tau), \quad \text{then}\quad (A,s) \sim (B,s+\tau) \qquad (3)$$

This means that, if an agent is indifferent ($\sim$) between commodity A, delivered at timepoint t, and commodity B, delivered at timepoint $t + \tau$, he will still be indifferent when A was delivered at a different timepoint s and B at timepoint $s + \tau$ (Strotz, 1955; Koopmans, 1960; Fishburn and Rubinstein, 1982). Indifference refers to the situation where a decision maker chooses all commodities with equal probability. It is assumed that the commodities have identical utility at the point of indifference. Stationarity implies that the indifference between two choice alternatives should depend only on the difference in the delays that the outcomes can be realised, and, given the time-difference between the outcomes remains the same, indifference should be preserved at all different timepoints. Hence, if both options were deferred into the future by the same time interval, subjects should still be indifferent between the options. For example, if you desire to receive \$10 in 5 days as much as receiving \$50 in 20 days, then you will still desire to receive \$10 in 15 days as much as receiving \$50 in 30 days, i.e., when both delays are prolonged by 10 days.

DUT posits that the discounting rate by which future commodities are devalued should be constant (Samuelson, 1937). Constant discounting means that a given time delay has the same relative impact on preferences and values, regardless of when it occurs. Constant discounting is necessary to warrant time-consistency, i.e., the postulate that ordering of preferences at a later timepoint should be identical to current or earlier preference orders (stationarity). Variable discount rates would hence be conflicting with the assumption of stationarity and consistency of preference orders (see below and Fig. 1 for details on why this is the case). Exponential discount functions have constant discount rates and were adopted by DUT (Samuelson, 1937; Lancaster, 1963; Fishburn and Rubinstein, 1982; Benzion et al., 1989; cf., Ainslie, 1975; cf. Prelec and Loewenstein, 1991; Fehr, 2002), for example (Lancaster, 1963):

$$(A,t) \sim A\,\mathrm{e}^{-k(t-t_0)} \qquad (4)$$

Eq. (4) states that the agent should be indifferent between a reward with the amount A, delivered at a future timepoint t, and the reward amount A at the present timepoint t_0, exponentially discounted for the interval $t - t_0$, with k being an individually different discount value. In other words, the expected utility of a future outcome can be expressed as an exponential decay function of the same outcome realised today.

One of the hallmarks of rational choice theories is the assumption that decision makers strive to maximise utility, and/or to optimise their cost–benefit function. In behavioural ecology, optimal foraging theory (Stephens and Krebs, 1986) makes similar assumptions about decision-making. Although coming from a different academic discipline, it shares many features with normative models in economics, and is therefore worthwhile discussing here. In biology, utility maximisation in the context of intertemporal choice can be viewed as maximising the Darwinian fitness, with, for example, energy intake per time unit as a proxy for Darwinian fitness[3] (Stephens and Krebs, 1986; Stephens and Anderson, 2001). This concept is called rate maximisation. In formal terms, it is assumed that organisms maximise, at least in the long run, the ratio of food intake and the time needed to obtain or consume the food, as described by the following quantity (Stephens and Krebs, 1986):

$$\max \frac{\sum_{i=1}^{\infty} G_i}{\sum_{i=1}^{\infty} t_i} \qquad (5)$$

where G_i represents the net energy gain obtained from consuming the ith food item (here basically corresponding to its amount), and t_i represents the time between food item i and the previous food item $i - 1$. In a choice between large, delayed and small, short-term rewards, rate maximisation predicts that animals prefer large rewards when the ratio of reward amount per time unit is higher for the large than for the small reward, e.g., when the animal chooses between 10 items of food in 6 s or 4 items in 3 s. If the waiting time for large rewards increased to, say, 9 s, rate maximisation would predict preference for the small reward.

3. Irrational intertemporal decisions: anomalies in intertemporal choice

3.1. Violation of the stationarity axiom

Stationarity predicts that the ranking of preferences between several future outcomes should be preserved when the choice outcomes are deferred into the future by a fixed interval. This has been investigated in an empirical study where human subjects chose between pairs of monetary rewards available after different delays (Green et al., 1994). Subjects preferred a small, short-delayed over a large, long-delayed reward. However, when the delays to both rewards were advanced by the same time interval, their preference reversed away from the small towards the large reward. Notably, the prolongation of the delays resulted in a preference reversal even though the difference in delays remained identical (Green et al., 1994). These time-inconsistent preferences represent a violation of stationarity which is sometimes called the common difference effect (Prelec and Loewenstein, 1991; Frederick et al., 2002). In its extreme form, a literal discontinuity of preference has been reported when immediate rewards are involved (immediacy effect; Thaler, 1981; Benzion et al., 1989). Numerous studies with human subjects (Ainslie, 1975; Thaler, 1981; Logue, 1988; Benzion et al., 1989; Loewenstein, 1992; Kirby and

[3] Of course, from a biological point of view, other aspects than merely energy intake are also of importance to an organism, such as reproduction and sleep, but for reasons of simplicity, we restrict this review to the discussion of energy intake.

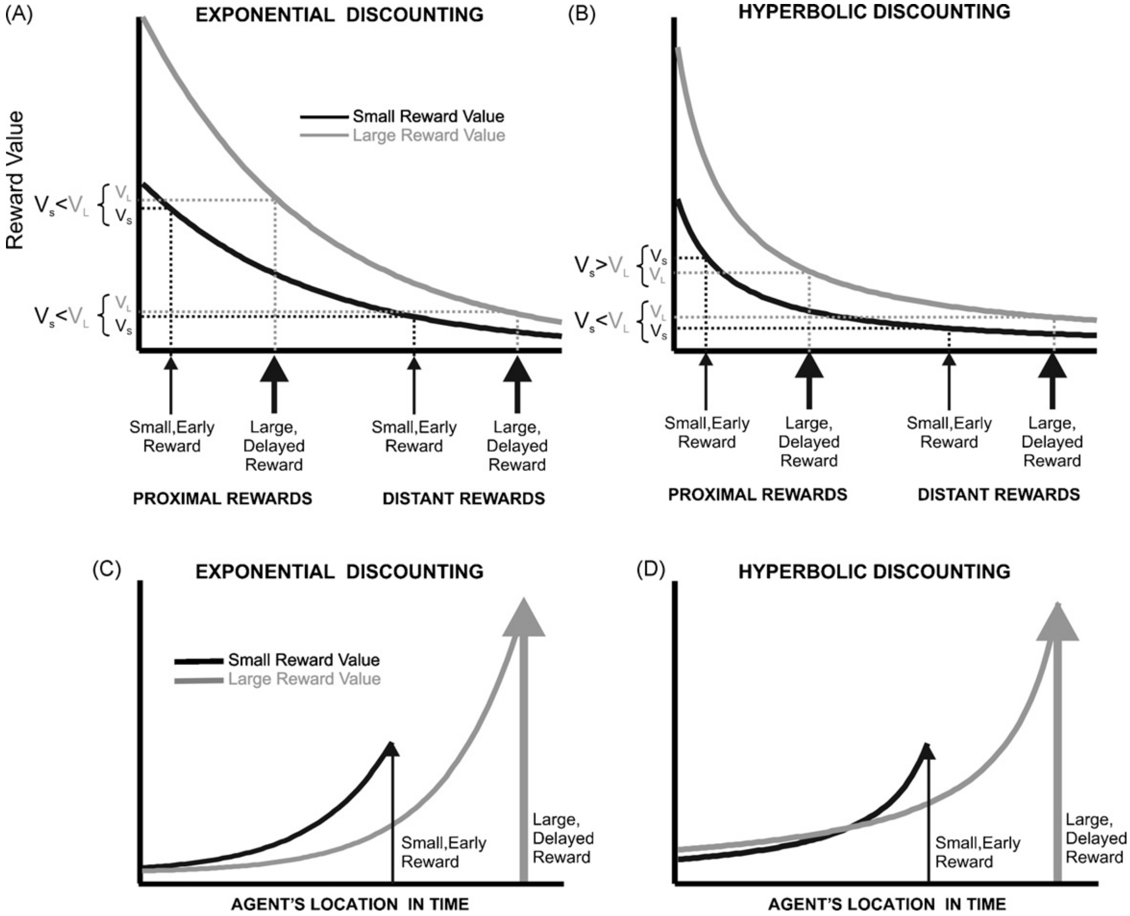

Fig. 1. Preference reversals can be better explained by hyperbolic than exponential discounting. The figure depicts the situation where a subject first chooses between a small, early and a large, delayed reward (proximal rewards), and subsequently, both rewards are deferred in time by the same time interval (distant rewards), thus preserving the delay-difference between them. The figure plots the discounted value of a future reward (y-axis) as a function of reward amount and delay. Grey lines represent the discounted value of the large reward, black lines the value of the small reward. (A) The x-axis depicts the delay to the reward, fat arrows indicate a large, delayed reward, slim arrows a small, early reward. Due to constant discounting in the exponential function, the value of the large, delayed reward V_L is larger than the value of the small, early reward V_S when both rewards are temporally proximal, and also when they are deferred by the same time interval, so that always holds: $V_S < V_L$. (B) In hyperbolic discounting, the values of large and small rewards reverse when the rewards are deferred into the future: whereas $V_S > V_L$ when both rewards are relatively temporally proximal, $V_S < V_L$ when they are relatively distant. Often, this rationale is illustrated in a somewhat different fashion, as shown in (C) and (D). The x-axis depicts the temporal distance to the forthcoming reward from the perspective of the agent looking into the future, the attractiveness of the rewards increases with decreasing temporal distance. The x-axis displays the agent's position in time with respect to the temporally proximal and distant rewards. Long, fat arrows indicate a large reward, short, thin arrows a small reward. (C) Exponential discounting: the small reward value is continuously higher than the large reward value for distant and proximal rewards. (D) Hyperbolic discounting: when both rewards are temporally distant (left of the intersection of the lines), the value of the large reward (grey line) exceeds that of the small reward (black line), and subjects should consequently prefer the large over the small reward. Due to the cross-over of the curves, the value of the small reward exceeds that of the large reward once the subject gets close in time to the small reward (right of the intersection), and they should now prefer the small over the large reward.

Herrnstein, 1995; Green et al., 1997; Frederick et al., 2002; McClure et al., 2004; Rohde, 2005), pigeons (Chung and Herrnstein, 1967; Rachlin and Green, 1972; Ainslie, 1974; Green et al., 1981) and rats (Ito and Asaki, 1982; Bennett, 2002) replicated and confirmed the common difference and immediacy effects.

3.2. Violation of the assumption of constant discounting

Stationarity implies constant discounting, i.e., a neutral attitude towards time delay: a given delay should have the same relative impact on utility regardless of when it occurs. Several authors (Ainslie, 1975; Rachlin et al., 1991; Green and Myerson, 1996; Frederick et al., 2002) pointed out that preference reversals as discussed above cannot be explained by constant discounting functions, such as exponential discounting. As theoretically suggested by Ainslie (1975), and later empirically shown by Mazur (Mazur, 1984, 1987, 1988; Grossbard and Mazur, 1986) and others (Thaler, 1981; Benzion et al., 1989; Rachlin et al., 1991; Myerson and Green, 1995; Green and Myerson, 1996; Rohde, 2005; Jones and Rachlin, 2006; Glimcher et al., 2007), rewards delivered with short delays are more steeply discounted than rewards with longer delays. Mathematical fits to the empirical data obtained in these studies indicated that a hyperbolic model:

$$V = \frac{A}{1 + kD} \tag{6}$$

approximates the data better than an exponential model, for example:

$$V = A\,\mathrm{e}^{-kD} \tag{7}$$

where in both equations, V represents the temporally discounted reward value (utility), A the amount of expected reward, D the delay between response and reward, and k an individually different discount rate.

Fig. 1 plots the value curves for two rewards differing in quantity and delay and shifted to the future by the same time interval. Fig. 1A shows the value curves for exponentially discounted rewards, Fig. 1B displays the curves for hyperbolically discounted curves.

In the exponential model, the value of the large reward V_L exceeds the value of the small reward V_S in both temporally proximal and distant reward situations because of the model's constant discount rate ($V_L > V_S$ always holds; cf. Fig. 1A). In contrast to this, the discount rates in the hyperbolic model are not constant over time. Instead, hyperbolic discounting is characterised by high discount rates over short horizons, but low discount rates over long horizons. This results in the reversal of the order of values, as illustrated in Fig. 1B: while the small reward value is higher than the large reward value ($V_S > V_L$) in the temporally proximal reward situation, V_S is smaller than V_L ($V_L > V_S$; cf. Fig. 1B) for distant rewards, although the time difference between both rewards is identical in the proximal and distant situation. Hence, it is difficult to explain preference reversals with a function assuming constant discount rates, such as the exponential model, but such reversals can be better accounted for by an asymmetric discount function, as for example the hyperbolic function in Eq. (6), a quasi-hyperbolic function (Laibson, 1997; a quasi-hyperbolic function approximates a hyperbolic function, but consists of often several, not necessarily hyperbolic subfunctions), or an exponential model similar to Eq. (7), but where the discounting factor k is variable and depends on the amount of the anticipated reward (Green et al., 1994, 1997). Despite some recent challenges (Schweighofer et al., 2006), the hyperbolic model is still the most widely accepted description of time preference.

On a side note, although the observations of intertemporal preference shifts and hyperbolic discounting represent violations of some of DUT's axioms, it may not necessarily indicate that reversing preferences over time is irrational per se. As argued by several authors, a re-evaluation of the current consumption plan and a subsequent change of course of action actually constitutes the more rational behaviour in many situations (Strotz, 1955; Kogut, 1990; Heath, 1995; Arkes and Ayton, 1999; Karlsson et al., 2005). For instance, according to most choice theories, rational decisions should be exclusively prospective, and not retrospective. Thus, a decision maker should choose the option with the highest expected (i.e., future) value. When evaluating the interim success of a project that the decision maker has previously committed to, earlier investments should play no role in the decision whether to continue the investment, or whether to abandon the project and invest in another one. The opposite behaviour, i.e., the escalation of commitment to a failing project although an alternative activity would promise better results, is called 'the sunk cost effect' (Kogut, 1990; Heath, 1995; Arkes and Ayton, 1999; Karlsson et al., 2005). Such escalation of commitment seems to be related to the amount of previous investments, and has been shown in humans (Kogut, 1990; Schaubroeck and Davis, 1994; Heath, 1995; Arkes and Ayton, 1999; Moon, 2001; Karlsson et al., 2005) and animals (Arkes and Ayton, 1999; Navarro and Fantino, 2005). It is unclear why subjects show the sunk cost effect, and several attempts have been made to explain the motivation for escalating commitment, including hesitation to waste resources (Arkes and Ayton, 1999), loss aversion (Schaubroeck and Davis, 1994), desire to complete a job (Moon, 2001) or project-based mental accounting (Heath, 1995). Despite the disagreement on the causes, it is generally undisputed that the sunk cost effect represents a deviation from rationality, because subjects consider the past in their decisions, therefore frequently and systematically prefer options with lower expected values, and even choose projects that bring about an accumulation of costs and losses. An example includes the continued investment of money, time and resources into the development of the supersonic plane 'Concorde', although the weak economic prospects of the plane were evident long before project completion.[4] This shows that perseverating on a once preferred choice option may lead to a less than optimal outcome in the long run. A rational decision maker, on the other hand, whose decision is only guided by weighing the expected values will change his preferences whenever options other than the initially preferred one yield higher expected values. Hence, time inconsistencies can, under some circumstances, be perfectly consistent with rational models of choice.

3.3. Violation of the assumption of utility maximisation

The rate maximisation hypothesis predicts that preferences and preference shifts should depend exclusively on the ratio of reward amount (or monetary value) and the duration between the rewards. In many situations, however, decisions are not determined by this ratio, but only by the waiting time preceding the rewards. In those cases, shifting the preference to the temporally proximal outcome is frequently not the optimal behaviour. As an example take an animal that chooses between a small, always immediate reward and a large, gradually delayed reward. Assume that the inter-trial interval is adjusted so that the total trial length is identical in all trials and independent of delay length and other factors. The rate maximisation hypothesis predicts that subjects should always choose the large reward, independent of the delay between response and reward, because only then would they maximise the total energy intake per trial, or per experimental session, respectively. However, neither pigeons (Rachlin and Green, 1972; Ainslie, 1974; Grossbard and Mazur, 1986; Mazur, 1988; Kalenscher et al., 2005b), nor rats (Evenden and Ryan, 1996;

[4] Because of this analogy, the sunk-cost effect is also frequently called 'Concorde fallacy' (Arkes and Ayton, 1999).

Cardinal et al., 2000; Winstanley et al., 2004, 2006; Roesch et al., 2006), mice (Isles et al., 2003, 2004), or monkeys (Hwang et al., 2006; Louie and Glimcher, 2006) show the predicted perseverance on the large reward alternative, but instead reverse their preference to the small, immediate reward once the large reward delay exceeds an individual threshold limit. The strong preference for short-term options suggests that animals satisfy a short-sighted decision rule (heuristic) which minimises waiting time, but does not necessarily yield the economically best outcome.

3.4. Gains and losses in intertemporal decision-making

DUT did not make a particular distinction between the treatment of losses and gains. Hence, the aversiveness of losses should be discounted as much as the attractiveness of gains when the outcome is more and more delayed. However, Thaler (1981) reports that the attractiveness of gains is reduced faster than the aversiveness of losses, implying a different discount rate for gains than for losses. This 'sign-effect' suggests a different mental treatment of gains and losses, and is therefore difficult to reconcile with DUT's assumption that intertemporal choice can be condensed into a single discount function. An even greater challenge for DUT is the observation that many human subjects prefer to expedite a loss instead of delaying it. If losses loom less when they are temporally remote, as predicted by DUT, subjects should be ready to defer losses into the future. However, many subjects actually prefer to incur a loss or an aversive event immediately rather than delay it (Benzion et al., 1989; Berns et al., 2006).

3.5. Further anomalies

DUT's postulate that the different motives in intertemporal decision-making can be condensed into a single discount function implies that discount rates should be the same for all goods and categories, and should be independent of the way a problem is mentally processed. This implication has been challenged by a range of other findings, in addition to the sign-effect discussed above. Framing, i.e., the way a problem is presented, has been shown to have an effect on the discount rate. For example, subjects who bought a good and expected its delivery in 1 year were willing to pay an extra of $54 dollars to have it delivered immediately. However, subjects who bought the same good, but expected its immediate delivery, demanded a compensation of $126 dollars if the delivery were delayed by 1 year (Loewenstein, 1988). Moreover, several studies have shown that large positive outcomes are discounted at a lower rate than small outcomes ('magnitude effect'; Thaler, 1981; Benzion et al., 1989; Green et al., 1994). Frederick et al. (2002) list a range of further anomalies. This list includes a preference for improving sequences, violations of independence, and preference for spread. Preference for improving sequences means that subjects prefer a stream of increasing comfort over a stream of decreasing comfort, for example, an increasing salary profile, even when the mean comfort is identical. As indicated above, violations of independence, and preference for spread refer to the fact that subjects prefer to spread the acquisition and/or consumption of commodities in time. For example, a person's current preference for pizza depends on whether she had pizza yesterday, or will have pizza tomorrow.

3.6. Alternative theories

In the past, it was mainly economists, psychologists and behavioural biologists who proposed alternative models of intertemporal choice to explain these anomalies. Early economic theories stressed the emotional and motivational impact of waiting for dated outcomes, for example, the pains and pleasures of anticipating an attractive reward, the discomfort of deferring immediate gratification, the displeasure of abstaining from consumption, etc. (see Loewenstein, 1992, for a historical overview). In this tradition, some more recent theories claim that the utility of a future commodity is derived not only from the actual consumption of the good, but also from the anticipation of this good (Loewenstein, 1987; see Section 4.6 for more details). Also somewhat inspired by the cognitive flavour of the early economic models, the temptation account by Gul and Pesendorfer (2001) posits that subjects experience disutility from not choosing the most tempting, i.e., immediate option. Subjects would hence be more content if the immediate option was not available, and they hence adopt measures to avoid future temptations. Yet another theory stresses that the total amount of discounting over a time span increases as the interval is more finely partitioned (Read, 2001). Such 'subadditive discounting' is an alternative to hyperbolic discounting, and predicts that the discount rates for, say, three 8-month intervals should be different to the discount rate across the sum of the three intervals, i.e., 24 months, but should not be different across the three 8-month intervals. Read (2001) found some empirical support for this idea. Other theories demanded a radically new view on intertemporal decision-making, and went so far as to assume that decisions can be best modelled by assuming non-stationarity of personal preferences, which is frequently portrayed using a somewhat catchy analogy of an 'intrapersonal conflict between multiple temporally situated selves' (Thaler and Shefrin, 1981; Laibson, 1997; Fudenberg and Levine, 2006; refer to Sections 4.7–4.9 for a detailed coverage of these models). 'Habit formation', 'reference point shifts' as predicted by prospect theory, and visceral influences on the decision have also been suggested to influence intertemporal decisions (see Frederick et al., 2002, for overview).

Yet another theory (Staddon and Cerutti, 2003) posits that hyperbolic discounting in animals is a by-product of temporal control and can be explained by the linear waiting rule which, in its simplest form, posits a linear relationship between response latency (in this case, the time until an animal makes a self-initiated response after having consumed a reward) and the delay between reward deliveries. Because the linear waiting rule appears to be obligatory, i.e., animals follow it even if they delay or even prevent reward delivery by doing so, the rule may represent a more fundamental temporal discrimination principle than it may seem at first glance. Staddon and Cerutti (2003) discussed the implications of linear waiting for a whole range of choice phenomena, hyperbolic discounting being one of them. Scalar expectancy theory by Gibbon (1977; cf. also Church and Meck, 2003) can also account for time inconsistencies. Scalar timing is the observation that the variance in accuracy of a time interval estimate is proportional to the length of the to-be-timed interval. Consider a subject who prefers a small, immediate over a large, delayed reward. When a fixed time interval is added to both delays, i.e., when both intervals become relatively longer, the distributions of the time estimates for both intervals will eventually overlap because of scalar timing, and discrimination between the intervals becomes imperfect. Because the *perceived* difference in time lengths is hence decreasing when both delays become equally longer, the relative impact of delay on the decision diminishes, the relative impact of reward amount increases, and preference eventually reverses. Moreover, it has been argued that good discrimination of time intervals yields higher long-term reward rate (Stephens, 2002; Stephens et al., 2004). Because discrimination of long intervals is imperfect, evolution may have favoured preference for short delays and, hence, the use of short-sighted decision rules.

A number of other attempts have been made in behavioural ecology and psychology to explain why non-human animals do not maximise reward rate. Many models presume that animals achieve optimality in choice situations through simple choice heuristics and rules-of-thumb, such as a win-stay-lose-shift strategy or linear waiting (Staddon, 2001; Staddon and Cerutti, 2003). Hence, animals do not really maximise, they simply follow a rule that happens to yield optimal results in ecologically valid life situations. For intertemporal choices, this means that, although the animals' delay-sensitive choice heuristics fail in artificial laboratory settings, they could produce rate maximisation in real life (Kacelnik, 1997; Stephens and Anderson, 2001; Stephens et al., 2004). Some researchers suggested that, rather than the usual binary choice requirement, a more ecologically valid scenario could correspond to a patch-like, multi-source situation which entails the decision whether to finish exploiting a current food site (short-term availability of small food resources), or abandon the food site preliminarily and travel to a next, potentially richer patch (long-term and less certain, but larger food quantity; Stephens and Anderson, 2001; Stephens et al., 2004). In several studies, Stephens and Anderson (2001; Stephens et al., 2004) have shown that short-term heuristics that aim to reduce waiting time indeed yield rate maximisation in such patch-like situations.

The multitude of different attempts to explain time inconsistencies presented in this section shows that intertemporal choice is multifaceted and can be approached from many different angles. The newly emerging field of neuroeconomics adds a new perspective to the discussion by requiring that a viable theory should have a biological basis. Although far from being able to explain all anomalies within a unified theory, the neurobiological approach provides new and promising insights into intertemporal choice.

3.7. Summary

There is evidence that DUT as a normative intertemporal decision theory has limited descriptive validity because it fails to adequately describe the reality of intertemporal choice behaviour. Common difference and immediacy effects, and the fact that preference reversals occur after deferring all choice alternatives into the future by the same time interval, violate assumptions of consistent choice, such as the stationarity axiom. Such preference reversals cannot be explained by models assuming a constant discount rate, whereas they can be accounted for by hyperbolic discount functions. Accordingly, in many studies, hyperbolic models consistently provided the better fits to the empirical behavioural data compared to exponential or linear models (Mazur, 1984; Grossbard and Mazur, 1986; Green et al., 1997; Glimcher et al., 2007). Hyperbolic discounting is therefore widely accepted in the literature as an appropriate description of reward devaluation over time. Furthermore, animals' preferences depend on the waiting time preceding the rewards, but not on the ratio of reward amount and time-between-rewards. Hence, inconsistent with the original optimal foraging formulation, animals seem to 'satisfice' (i.e., satisfy and suffice, in this case, satisfy a myopic decision rule), rather than maximise (i.e., maximise the energy intake rate). Other anomalies, such as the sign effect (different treatment of gains and losses) and magnitude and framing effects challenge the view that intertemporal choice can be condensed into a single discount function.

In conclusion, behavioural evidence suggests that humans and non-human animals systematically violate many of the crucial assumptions of DUT when making intertemporal decisions. A number of alternative models have been proposed to account for these violations, some of which will be discussed in greater detail later on in Section 4.

Note the similarities in the anomalies in DUT and EUT. The sign effect in DUT, for example, corresponds to the reflection effect in EUT (risk aversion in the domain of gains, but risk seeking in the domain of losses), the common difference and immediacy effects correspond to the common ratio and certainty effects in EUT (overweighting of certain outcomes), and the implications of both theories are challenged by framing effects (see Kahneman and Tversky, 1979 for anomalies in EUT, and Prelec and Loewenstein, 1991, for a comparison of anomalies in DUT and EUT).

As a side note, normative models, such as DUT, are not descriptive theories, and are not meant to be, although they are frequently accepted as such. Hence, empirical findings contradictory to the models' assumptions and predictions are not necessarily a challenge to the models per se, but merely illustrate how real decision makers deviate from the ideal optimal decision maker assumed by the models. Interestingly, though, subjects often do not regard their own violations of DUT axioms as irrational when pointed out to them. This is different to violations of EUT (cf. Frederick et al., 2002). This observation does not squelch the doubts about the descriptive validity of DUT, but it certainly calls the intuitive validity of DUT into question.

4. The neuroscience of intertemporal choices

4.1. Intertemporal decision-making—a challenge for the cognitive neurosciences

The challenge in neuroscientific research lies in discovering the neural correlates of the cognitive mechanisms underlying intertemporal choices, and, more importantly, formulating neural models capable of explaining why subjects violate the axioms and assumptions of DUT. With this goal in mind, we will address the following issues in the remainder of this article:

1. Identify the networks involved in representing the two essential decision parameters 'delay' and 'reward amount', and reveal the mechanism that weighs these parameters and converts them into a categorical response.
2. Investigate neural circuits involved in representing subjective value in the brain. Determine whether the choice parameters are neurally integrated into a single representation of the discounted value signal ('common currency'; cf., Montague and Berns, 2002), or whether distributed choice mechanisms account for intertemporal choice.
3. Understanding non-constant, e.g., hyperbolic, discounting is essential for understanding time inconsistencies. The most essential challenge in the cognitive neurosciences is to identify the neural mechanisms that produces non-constant discounting and time-inconsistencies.

4.2. Neural representation of the decision variables 'reward delay' and 'reward amount'

An intertemporal decision is determined by the values of the different outcomes, and the delay until the outcomes can be realised, or, in other words, by the integration of expected reward amount and time-to-reward. Both constituents, i.e., reward amount and time, are represented in the brain.

Neurons in many different brain areas represent information that is pertinent to the amount of an actual or expected reward. The activity of single units in dorsolateral prefrontal cortex (DLPFC; Leon and Shadlen, 1999; Wallis and Miller, 2003), orbitofrontal cortex (OFC; Wallis and Miller, 2003; Roesch and Olson, 2004; Van Duuren et al., 2007), and post-arcuate premotor cortex (Roesch and Olson, 2004) of rats and monkeys, and in the avian equivalent of the prefrontal cortex (Kalenscher et al., 2005b) is modulated by the magnitude of an expected reward, either during the presentation of a reward-predicting cue, or during the delay between the cue and the delivery of the reward, hence before the actual reward is delivered. This has been taken as evidence that frontal structures play a role in maintaining reward value in memory while waiting for the reward, representing ongoing events, and monitoring the expected consequences of choices (Schoenbaum et al., 1998; Montague and Berns, 2002; Winstanley et al., 2004, 2005; Van Duuren et al., 2007). Likewise, single unit activity in the striatum, in particular in the ventral striatum, including nucleus accumbens (NAc), correlates with the amount of an expected reward in monkeys (Hollerman et al., 1998; Hassani et al., 2001; Cromwell and Schultz, 2003) and birds (Izawa et al., 2005). In addition, neurons in monkey and rat OFC and lateral prefrontal cortex (Schoenbaum et al., 1998; Tremblay and Schultz, 1999; Hikosaka and Watanabe, 2000; Wallis and Miller, 2003; Roesch and Olson, 2004; Padoa-Schioppa and Assad, 2006; Roesch et al., 2006), DLPFC (Watanabe, 1996; Wallis and Miller, 2003), basolateral amygdala (BLA; Schoenbaum et al., 1998; Baxter and Murray, 2002), and NAc of monkeys (Hassani et al., 2001) and birds (Yanagihara et al., 2001) discriminate between expected rewards that differ not only in their quantity, but also in type and quality. OFC neurons in monkeys show anticipatory and reward discriminating activity even before the instructive cue occurs (Hikosaka and Watanabe, 2004), hence suggesting that these neurons reflect long-range reward expectancy. Reward-discriminating units have also been found in a range of other structures, most notably in the midbrain ventral tegmental area and substantia nigra pars compacta (Schultz, 2002, 2004; Tobler et al., 2005), but also in lateral intraparietal area, anterior and posterior cingulate cortex, hypothalamus and others (Platt and Glimcher, 1999; cf. Schultz, 2002, 2004; McCoy and Platt, 2005). Neurons in prefrontal cortex (PFC) and parietal cortex also track the performance- and reward history in an oculomotor task (Hasegawa et al., 2000; Sugrue et al., 2004), and statistically predict future performance (Hasegawa et al., 2000; see also Roelfsema, 2002).

In addition to reward amount, time is likewise processed in various areas in the brain. Human research indicates that several distributed brain regions play a role in processing interval timing, including striatum, cerebellum, thalamus and various parts of the cortex (for overviews, cf., Ivry, 1996, 1997; Matell and Meck, 2000; Buonomano and Karmarkar, 2002; Ivry et al., 2002; Durstewitz, 2004; Buhusi and Meck, 2005). Such timing could be implemented either by a central clock or pacemaker functioning as the brain's metronome (cf. Buonomano and Karmarkar, 2002; Buhusi and Meck, 2005). Alternatively, newer models conjecture that interval-timing is implemented by so-called climbing activity, i.e., the gradual increase of neural discharge rate across a delay. It has been shown that the slope and the timepoint of maximal activation of this climbing function is scaled to the duration of the to-be-timed interval. Given the existence of a read-out mechanism once the ramping activity reaches a threshold, interval timing can be accomplished by adjusting the slope of the climbing function to the required delay length (Durstewitz, 2003, 2004). Such interval-timing-dependent ramping activity has been found in posterior thalamus (Komura et al., 2001), posterior parietal cortex (Leon and Shadlen, 2003; Janssen and Shadlen, 2005), inferotemporal cortex (Reutimann et al., 2004), DLPFC (Kojima and Goldman-Rakic, 1982; Rainer and Miller, 2002; Brody et al., 2003; Sakurai et al., 2004) and its equivalent structure in the avian brain (Kalenscher et al., 2006b), cingulate cortex (Kojima and Goldman-Rakic, 1982), ventral striatum (Izawa et al., 2005), primary visual cortex (Shuler and Bear, 2006), frontal and supplementary eye fields (Sato and Schall, 2003; Roesch and Olson, 2005a), and premotor and supplementary motor cortex (Roesch and Olson, 2005a).

Staddon and Cerutti (2003) mathematically proved that linear waiting as an obligatory interval time discrimination rule can account for a range of timing and choice phenomena, such as Weber's law of scalar timing (cf. Church and Meck, 2003), difference in risk attitude in the domains of gains and losses (Kahneman and Tversky, 1979), and also hyperbolic discounting. Thus, if the neural processes examined in this section are causal mechanisms of interval timing, these mechanisms could be sufficient to explain hyperbolic discounting and hence many of the anomalies discussed in Section 3. This would eliminate the need to look for the elusive neural counterpart of intertemporal choice. However, because evidence that climbing functions can be explained within the linear waiting framework is still elusive, this is just speculative at this point.

Note that the majority of the cited studies on interval timing involved time delays in the range of a few seconds. These time intervals match the intervals used in many of the animal choice experiments, but not necessarily those used in human research. In addition to the fact that human subjects are usually instructed to imagine the delays, which means they do not experience the ends of the delays, the interval lengths are typically in the range of months and years, not in seconds. The measurement of interval durations in a long and short range may recruit entirely different neural mechanisms (cf. Hinton and Meck, 1997; Lewis and Miall, 2003). It is hence uncertain whether the timing mechanisms discussed in this section actually play a role in intertemporal choice when imagined delay intervals of well above several seconds are involved.

4.3. Neuroanatomy of intertemporal decisions

Much of our knowledge about the neuroanatomy of intertemporal decision-making stems from neuropsychological research with patients that show symptoms of abnormally disadvantageous delay-discounting, e.g., future-blindness and exaggerated impulsiveness (Bechara et al., 1996, 1998, 2000a, b; Hartje and Poeck, 1997; Kolb and Whishaw, 2003). Such pathologies include attention deficit hyperactivity disorder, drug addiction, problem gambling, and frontal lobe syndrome. All of these conditions presumably involve a pathological modulation of frontal lobe function. The PFC is generally considered the crucial bridge in the perception-action cycle that mediates action-reward contingencies across time (Quintana and Fuster, 1999; Fuster, 2000). Research on frontal lobe dysfunction has revealed that patients with lesions in their ventromedial prefrontal cortex (VMPFC) tend to overly strongly discount, or even neglect, the future consequences of their decisions, be they appetitive or aversive (Bechara et al., 1996, 1998, 2000a, b). Due to this evidence, and the fact that frontal cortex is generally associated with decision-making (Lee et al., 2007), prefrontal regions are generally considered prime candidate structures to control delay discounting and impulsiveness (in this context defined as the propensity to choose the small, immediate reward).

Recently, animal lesion studies have somewhat challenged this conclusion. One study showed that lesions of the core of the rodent NAc, but not the medial PFC (mPFC) or the anterior cingulate cortex (ACC), resulted in a reduced delay tolerance and increased impulsiveness in a delay discounting procedure (Cardinal et al., 2001). This points towards a special role of the ventral striatum, which contains the NAc, in producing time preference. There is also evidence for the involvement of the avian equivalent of the NAc in impulsive decision-making (Izawa et al., 2003, 2005), and recent neuroimaging experiments with human subjects suggested that striatal activation extending to the NAc is linked to choices of immediate rewards (Wittmann et al., 2007), and that the individual preference for immediate over delayed rewards covaries with differences in ventral striatum activations to reward feedback (Hariri et al., 2006). However, we also note that the core of the NAc is heavily dependent in its general functioning on prefrontal input (Voorn et al., 2004), which brings up the question to what extent the NAc's special role is implemented independently from, or in concert with, some prefrontal regions, e.g., OFC or agranular insular cortex.

Moreover, several studies have shown that the integrity of the BLA, and its connection with OFC, is necessary to flexibly adapt the representation of changing reward values and their link with reward-predicting cues during performance in reinforcer devaluation tasks, in which rewards are devalued through, for example, satiation, or conditioned taste aversion (Baxter et al., 2000; Baxter and Murray, 2002; Pickens et al., 2003). Delay discounting tasks likewise require the constant re-evaluation and updating of reward value representations during varying delays, indicating that amygdala–OFC interactions may be important for mediating intertemporal decisions, too. However, results are ambiguous: Although lesions of the rat BLA increased preference for short-term rewards lesions of the OFC and the subthalamic nucleus (STN) actually decreased impulsive choices in some studies (Kheramin et al., 2002; Winstanley et al., 2004, 2005), but increased impulsive choices in other studies (Mobini et al., 2002; Rudebeck et al., 2006).

Although the discrepant results in the OFC lesion studies could be also explained by subtle differences in task requirements or the spatial extent of the lesion, these studies nevertheless suggest that NAc, BLA, STN, and OFC appear to play different roles in impulsive choice behaviour. Winstanley and colleagues conclude that NAc and BLA may be important for representing and maintaining the subjective reward value across the delay, STN may be relevant for permitting basic Pavlovian associations, and OFC may play a role in monitoring and updating representations of expected rewards. Therefore, lesions of NAc and BLA should increase impulsiveness by impairing the representation or maintenance of incentive salience, but OFC lesions should induce perseveration by impairing the updating of subjective reward values during increasing delays. However, the functions of prefrontal subregions remain somewhat unclear, because a recent microdialysis study found evidence for dissociable roles of rat mPFC and OFC in impulsive decision-making, as reflected by differences in the monoaminergic regulation of these areas (Winstanley et al., 2006). The results of this study suggest that mPFC was relevant for representing the response-reward contingencies, and OFC played a direct role during decision-making, hence pointing to a direct involvement of rat OFC in choice behaviour that extends beyond simple outcome monitoring and representation. A recent electrophysiological study in rats (Roesch et al., 2006) that will be discussed in greater detail below, indicates that the rat OFC may contain dissociable, separate reward-processing networks that deal with the parameters delay and reward amount independently. This independent processing may explain some of the heterogeneity of the results of the OFC lesion studies. The role of other structures implicated in reward processing, processing of decision costs, decision-making in general and/or time estimation have also been examined in the context of intertemporal decisions, such as the shell of the NAc, anterior cingulate cortex, subparts of the rat mPFC including prelimbic and infralimbic cortex, the hippocampus, and the insula. Wittmann et al. (2007) found evidence that posterior insula was activated when human subjects choose delayed over immediate reward, suggesting that the insula may be involved in delaying gratification. The precise part of other structures in intertemporal choice is, at best, unclear (see Cardinal, 2006 for review). In summary, although their exact function remains to be resolved, evidence suggests the involvement of various limbic reward-related structures and frontal areas, including the OFC, in intertemporal decision-making.

4.4. Neural integration of 'reward delay' and 'reward amount'

In Section 4.2, we have outlined that the two essential parameters for time discounting, reward amount and interval time are represented in many different parts in the brain. In this section, we will address the question how, and if, these ingredients are neurally integrated to represent a discounted reward value. Following the evidence discussed in Section 4.3, a first guess for a convergence site is the OFC. To address whether OFC neurons indeed represent a compound of reward amount and delay, Roesch and Olson (2005b) trained monkeys in two tasks, one with a variable delay and a fixed reward amount, and another one with a variable reward, but fixed delay. OFC cells responded more strongly to a cue predicting the quantity of the upcoming reward when the monkeys expected a large compared to a small reward, and the same neurons were also more active to delay-predicting cues when the animals anticipated a short *versus* a long delay between cue and response. This suggests that reward proximity and quantity are processed by the same orbitofrontal neurons, implying they may be integrated on a single-cell level. However, delay and reward amount were not varied simultaneously in this study, hence it remains elusive whether the neurons were merely tuned to the respective task requirement, or whether they encoded the genuine, temporally discounted reward value. There is further preliminary evidence that units in the monkey lateral intraparietal area (Louie and Glimcher, 2006), monkey DLPFC (Hwang et al., 2006), rat ventral tegmental area (Roesch, personal communication) and possibly also human ventral striatum and amygdala (Gregorios-Pippas et al., 2005) represent, or even integrate, both decision parameters when choosing between differently dated and sized rewards.

One study provided direct evidence for an integration of reward proximity and amount on a single-cell level during

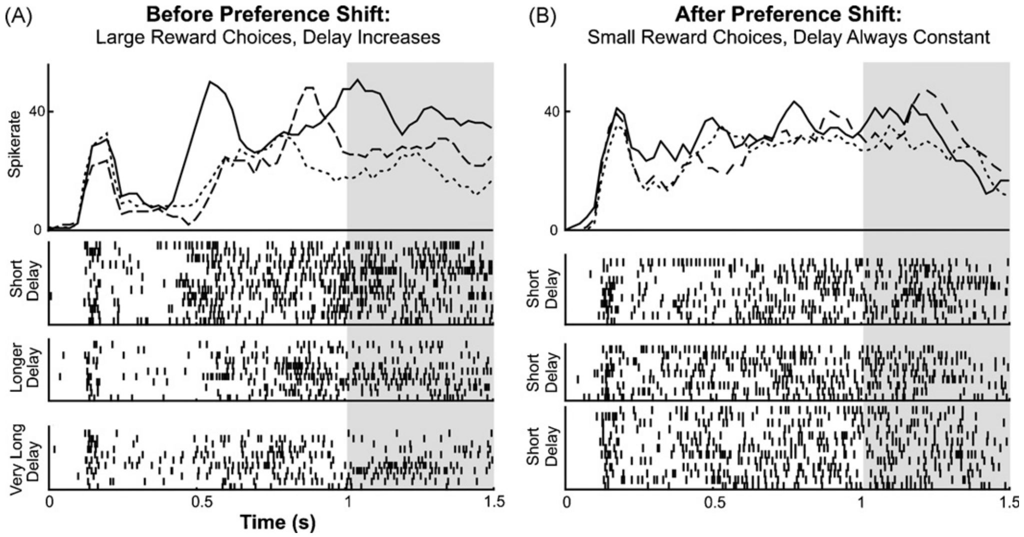

Fig. 2. Neural correlates of temporally discounted subjective reward value in the pigeon brain. Neural activity before and after the preference shift from the large, delayed to the small, immediate reward (delay onset at time zero). This figure shows the averaged and smoothed peri-stimulus time histograms (PSTHs) and raster plots of the reward-preceding sustained delay activity of one exemplar neuron. The PSTHs indicate the neuron's discharge rate (in Hz), each vertical bar in the raster plots represents one spike, each row corresponds to one trial. (A) Before the preference shift, pigeons preferred the large reward, but the delay increased across blocks of trials. The solid black line indicates the PSTH for early trials where the delay was minimal, the dashed line indicates the PSTH in large-reward trials where the delay was somewhat longer, and the dotted line indicates the PSTH in large-reward trials where the delay was maximal, and the preference shift to the small, always immediate reward was just about to occur. This figure shows that the final level of sustained activity (grey box) decreased as the length of the delay increased. (B) The PSTHs and raster plots of the same neuron across trials following the preference shift where the pigeon preferred the small, always immediate reward. There was no systematic variation in the neuron's discharge rate across the trials. In parts modified from Kalenscher et al. (2005b). Copyright 2005 with friendly permission from Elsevier.

intertemporal decision-making (Kalenscher et al., 2005b). Pigeons were trained in a delay discounting task to choose between a small, always immediate and a large, gradually delayed reward. The trial length was identical across all trials. The animals showed the typical within-session preference shift once the delay preceding the large reward exceeded an individually different tolerance limit, hence replicating the previous finding that subjects do not maximise their rate of energy intake, but rather employ a waiting-time sensitive choice heuristic. A neural correlate of this heuristic should likewise show delay-sensitivity.

Single-cell recordings in the nidopallium caudolaterale (NCL), a structure in the avian brain functionally comparable to the mammalian PFC (Mogensen and Divac, 1982, 1993; Kröner and Güntürkün, 1999; Diekamp et al., 2002; Kalenscher et al., 2003, 2005a; Lissek and Güntürkün, 2003; Reiner et al., 2004; Güntürkün, 2005; Jarvis et al., 2005) provided evidence for such a neural correlate. Some neurons showed significantly enhanced sustained delay activity between response and reward delivery (Figs. 2 and 3). The activity of a subset of these neurons was correlated with the temporal proximity to the expected reward, i.e., given a fixed reward amount, the activation magnitude decreased with increasing delay, and was unmodulated across trials with equal delays. Importantly, when comparing the activity levels of these neurons across trials with identical delay lengths, but different reward amounts, the neural activation magnitude was higher when the pigeons expected a large compared to a small reward. This indicates that the neurons' activity levels were a function of both decision parameters 'reward proximity' and 'quantity', and therefore presumably reflected the temporally discounted utility of the reward. In support of this conclusion, the compound neural activation level correlated with the pigeons' differential preference for the large or the small reward, and thus with the occurrence of the preference shift. Hence, this study suggests that the discounted reward value is represented on a single-neuron level.

If one attempts to discover a good indication of processes related to the irrationality of intertemporal choice patterns, it is necessary to show that the neural correlate follows the predictions of non-constant, disproportionate time discounting models. A fit of a linear, an exponential, and a hyperbolic model to the data of Kalenscher et al. (2005b) study showed that the hyperbolic and the exponential model approximated the data better than the linear model (Fig. 3). However, although the fit of the hyperbolic model was better than that of the exponential model, as determined by a goodness-of-fit criterion, this superiority was subtle, and future research needs to determine which function ultimately provides the better fit.

4.5. Is the discounted reward value represented on a single-cell or population level?

Not all electrophysiological studies are consistent with this finding. Two recent studies showed evidence for a distributed representation of intertemporal choice parameters. One study with chicks revealed that reward amount and delay were discretely coded in different subpopulations of neurons in the avian ventral striatum (Izawa et al., 2005). Another study (Roesch et al., 2006) found similar results for OFC neurons. In this experiment, rats were trained to respond to one of two fluid wells with initially identical delays and reward amounts. The authors recorded the activity of single OFC neurons, and varied either delay or amount independently, which allowed them to disentangle a reward-quantity from a delay-signal in the brain. They found a number of neurons that were active in anticipation of reward until shortly after reward delivery. In contrast to the firing properties of monkey OFC neurons described above, there was no evidence that single OFC cells in rats co-

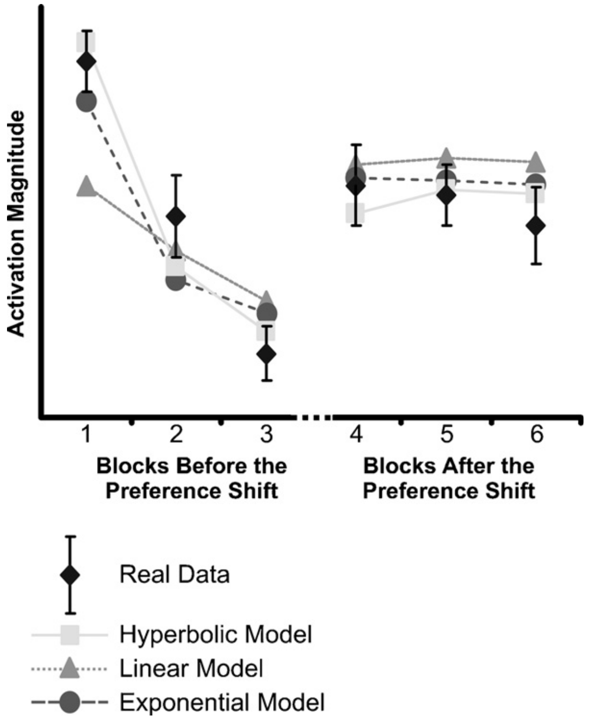

Fig. 3. Neural Activation pattern in the study by Kalenscher et al. (2005b) (see Fig. 2), and model fits. The black solid diamonds indicate the mean (±S.E.M.) activation magnitude, averaged across the neurons of interest. The left part of the panel (tick marks 1–3 on the *x*-axis, the tick marks are referring to block numbers) shows the development of the neural activity across trials before the preference shift, i.e., pigeons prefer the large reward and the delay length is increasing. The right part of the panel (tick marks 4–6) shows the activation pattern across trials following the preference shift, i.e., due to the long waiting time preceding large rewards, pigeons prefer the small reward and the delay length is constantly short. This figure confirms that the neural activity was negatively correlated with delay duration, but it also shows that the same neurons had higher activity in anticipation of large compared to small rewards when the delay preceding both rewards was equivalent (tick mark 1 compared to tick marks 4, 5, or 6). The other symbols show the model fits (light grey rectangles and solid light grey line: hyperbolic fit, grey triangles and dotted line: linear fit, dark grey circles and dashed dark grey line: exponential fit). The linear model approximated the empirical data clearly worse than the two other models. Note that the linear model could have approximated the neural data preceding the preference shift relatively well, but the fit was poor because of the inclusion of the data following the preference shift. Although the differences between the hyperbolic and the exponential model were subtle, a goodness-of-fit criterion confirmed that the hyperbolic function provided a better fit. Reprinted and in parts modified from Kalenscher et al. (2005b). Copyright 2005 with friendly permission from Elsevier.

represented amount and delay, as the activity of a subset of neurons correlated with amount, but not delay, and the activity of a different subset (negatively) correlated with delay, but not amount. This indicates that OFC neurons had dissociable representations of the values of differently delayed and sized rewards, suggesting that the discounted reward values were not represented on a single cell level, but on a population level.

Roesch et al. (2006, 2007) interpreted this result by proposing that OFC fulfils a double function: maintaining a reward-expectancy signal across a delay, and flexibly discounting the representation of increasingly delayed rewards independently of value representation. Their studies and others (Tremblay and Schultz, 1999; Montague and Berns, 2002; Kalenscher et al., 2005b, 2006a, b, see above) showed that a subset of neurons in OFC or related structures exhibit sustained activity across the delay preceding a reward. Such a signal bridges the gap between reward-predicting cues and reward delivery, and could thus facilitate the formation of cue-outcome or cue-value representations when reward delivery is delayed, for example in BLA (Saddoris et al., 2005) or NAc (Cardinal et al., 2001).

This dual task hypothesis may explain the conflicting results in the OFC lesion studies reported above. Animals that have been fully trained in a delay-discounting task have already formed strong associations between cue and delayed reward. Lesioning the OFC in those animals will deprive them of a discounting signal, and hence weaken their ability to flexibly update the discounted reward value. As a result, compared to controls, the lesioned animals will be more likely to wait for an increasingly delayed reward, hence showing a reduction in impulsivity, as reported by Kheramin et al. (2002) or Winstanley et al. (2004). On the other hand, animals lesioned before proper training may be unable to maintain the reward-expectancy signal across the delay, therefore having deficits in acquiring the cue-outcome association when the outcome is delayed. As a result, compared to control animals, their cue-outcome associations should become weaker with increasing delay lengths, and they should appear more impulsive, i.e., be less likely to choose increasingly delayed rewards, as has been shown by Mobini et al. (2002).

It is still unclear, though, why some evidence suggests that reward value is neurally represented on single-cell level (Kalenscher et al., 2005b; Roesch and Olson, 2005b), whereas other evidence (Izawa et al., 2005; Roesch et al., 2006) points towards a more complex, probably distributed encoding of time-discounted value. The dissimilarity of the results may have diverse reasons. First, they may be attributable to species-specific differences, implying that pigeons (Kalenscher et al., 2005b) and monkeys (Roesch and Olson, 2005b) may employ different discounting mechanisms than rats (Roesch et al., 2006). However, there is evidence that delay discounting is an evolutionary ancient process (Hodges and Wolf, 1981). It is therefore not unlikely that both species share similar discounting mechanisms. Second, the two parameters 'reward amount' and 'delay' were varied independently in the rat-study by Roesch et al. (2006), but combined in the pigeon-study by Kalenscher et al. (2005b). Thus, unlike the pigeons, the rats were never exposed to a situation that required them to integrate reward amount and delay. If OFC neurons are tuned differentially to different task requirements, it is not surprising that they reflected time and delay independently in the experiment by Roesch et al. (2006), but integrated both parameters in Kalenscher et al. (2005b) study. However, in the monkey study (Roesch and Olson, 2005b), delay and reward quantity were also not combined, but the OFC neurons nevertheless showed evidence of co-representing both parameters. These considerations imply that the results from either or all experiments may be peculiar, non-generalisable effects that are only valid within the particular task configuration. Third, although the NCL is believed to be equivalent to the PFC, the NCL-subdivision where the reported neurons were found may not be comparable to OFC, but may correspond to another part of the mammalian brain. Consequently, an integrated signal may, thus, be found elsewhere in the mammalian brain, for instance in more dorsolateral prefrontal (Hwang et al., 2006) or parietal regions (Louie and Glimcher, 2006), or in dopaminergic midbrain structures (cf. Tobler et al., 2005; Roesch, personal communication).

4.6. *Utility from anticipation—the value of passage of time in the brain*

As briefly mentioned in Section 3.6, the passage of time before an event occurs may have a utility in itself. According to the 'utility from anticipation' model (Loewenstein, 1987), the final utility of a future outcome is a combination of the utility derived from anticipating the outcome, and the discounted utility of the future consumption. For example, an exponentially discounted utility function may be deformed by an additional anticipatory-utility term, so that the resulting curve deviates from its original exponential shape. This theory can explain the sign-effect holding that temporal prospects in the domain of losses are discounted more strongly than in the domain of gains: if dreading future aversive outcomes is stronger than relishing appetitive outcomes, then waiting time would have a stronger impact on the overall disutility in the domain of costs/aversive events than on the overall utility in the domain of gains. It can also explain why subjects frequently speed up the delivery of aversive events instead of delaying it (Benzion et al., 1989, see Section 3.4) because they aim to end the disutility of dreading the looming detriment. Furthermore, it can also account for why the discount rate is not the same for all goods and categories of decisions.

Evidence for the theory of utility from anticipation comes from studying the neuroscience of dated aversive events (Berns et al., 2006). In a neuroimaging study, human subjects chose between pairs of electric shocks of different intensities, delivered at different timepoints after the choice. As shown before, subjects tended to speed up the shock delivery, i.e., when the voltages were identical, they generally preferred the shorter delay. Some subjects ('high dreaders') even preferred higher over lower shock intensities if the stronger shocks were administered after a shorter delay. The behavioural difference between 'high dreaders' and the other subjects ('mild dreaders') was unlikely due to a higher sensitivity to the noxious stimulation because, at the timepoints of shock administration, there was no or little difference in the effect of voltage or delay on the blood-oxygen-level dependent (BOLD) signal between the subjects in structures usually associated with somatosensory, visceral and emotional aspects of pain, the so-called pain matrix (including primary and secondary somatosensory cortex, anterior and posterior insula, caudal ACC, amygdala and other regions; see Fig. 4A). However, compared to the mild dreaders, the high dreaders had an earlier and more sustained BOLD response during the delay *before* the shock, i.e., in the anticipation phase, in several caudal regions of the pain matrix, including caudal ACC, posterior insula, and, most pronounced, in secondary somatosensory cortex (Fig. 4A). Moreover, the BOLD responses in secondary somatosensory cortex could be better approximated by a discount function incorporating a dread term than a discount function without such a parameter. Hence, the neural activations in caudal parts of the pain matrix tracked the between-subject differences in susceptibility to delay. Because these regions have been previously implicated in effects of attention, this study suggests that the experience of dread, and the motivation to speed up shock administration, comes from the attention devoted to the anticipated shock.

4.7. A single valuation mechanism or multiple choice systems?

Despite showing neural evidence for the (dis)utility of anticipation, Berns et al. (2006) did not address how and where the anticipatory (dis)utility was integrated with the discounted (dis)utility of the outcome. Moreover, this study was about dread, and it is unclear if and how the neural results translate to the domain of gains. In other words, it is still elusive how a non-constant, hyperbolic value-curve for positive outcomes comes about, and how this mechanism is implemented in the brain.

Roughly speaking, there are two schools of thought on this issue: according to one, genuine hyperbolic discounting is the key to understand time inconsistencies and other anomalies in intertemporal choice. That is, hyperbolic discounting in the domain of rewards or other appetitive events could stem from a single valuation mechanism for economic decisions (Samuelson, 1937; Rachlin, 2000; Montague and Berns, 2002; Roesch and Olson, 2004, 2005b; Kalenscher et al., 2005b; Padoa-Schioppa and Assad, 2006). Along with this idea, non-constant discounting could be driven, for example, by a non-linear integration of waiting time and reward amount on a single cell level (Kalenscher et al., 2005b).

Alternatively, preference reversals can be explained by positing a conflict between one's today's preferences when the outcomes of the current decisions are far away in time, and the preferences that will be held in the future when the outcomes are close in time. It is as if an agent's current 'self' exhibits different preferences than his future 'self' (non-stationarity of preferences). Several authors have therefore proposed that the processes resulting in this 'intrapersonal dynamic conflict' can be modelled by positing multiple economic 'selves'[5] in time (e.g., Thaler and Shefrin, 1981; Laibson, 1997; Fudenberg and Levine, 2006): there would be two 'selves' within one person, a myopic and a far-sighted 'self', who alternately take control over behaviour. Although not every author explicitly referred to temporally situated 'selves', many made comparable assumptions, and posited the existence of separate, competing decision processes, for example a 'hot' emotional system with a short time horizon, dealing for instance with the emotional temptation of short-term goals, the discomfort of deferring a proximate goal, or the impatience to realise a goal, *versus* a 'cool' far-sighted reasoning system involved in economic planning and cost–benefit trade-offs (e.g., Metcalfe and Mischel, 1999; Loewenstein and O'Donoghue, 2004; cf. also Sanfey et al., 2006). Although some theories (Fudenberg and Levine, 2006) assume that the different choice systems share the same base preferences, and differ only in how they regard the future, other models (Metcalfe and Mischel, 1999) propose that the different processes have at least partly different preferences, and drive the decision maker in opposite directions of choice. In addition, the models disagree in their assumptions on the strategies people use to regulate their own future behaviour, i.e., how the long-range or 'reasoning' system controls the short-range or 'emotional' system. For example, they contain different notions of self-control that impose different costs to the various sub-systems, for example, by reducing the future decision space through commitment to a choice option.[6]

Despite critical differences, all models have in common that temporal inconsistencies in preference arise from the specific interaction between the multiple choice systems. For example, because the impact of the 'hot' system is strong for short delays, but weak for long delays, people should discount rewards stronger at shorter delays, but less strong at long delays, and hence show temporal inconsistencies in preference.

4.8. Evidence for multiple decision networks in the brain

Inspired by the multiple-process models, Laibson (1997) proposed a quasi-hyperbolic utility function which is composed of two components, a β- and a δ-component. The δ-component corresponds to the economic planning system with exponential

[5] The terminology of 'multiple selves' is of course used in a metaphorical sense, the models imply that the behaviour of a single agent is determined by the interaction of two or more subsystems.

[6] Commitment usually refers to measures taken by an agent to avoid anticipated preference reversals by forcing himself to choose in the future the option that is currently preferred, for example, through restricting the future choice space. Restricting the future choice space can be accomplished by, for example, eliminating the option that seems inferior now, but may be tempting in the future (cf. Rachlin and Green, 1972; Ainslie, 1975; Laibson, 1997). As an example, take Odysseus who tied himself to the mast of his ship to avoid succumbing to his anticipated temptation to follow the sirens' call and therefore doom his ship. The issue of self-control and commitment is an entire field of research in itself, and although closely related to this article's topic, not further covered here.

discounting and a long time horizon, the β-component represents the short-term system, which gives extra weight to instant rewards. Neuroscientific evidence for the β–δ-model has been found in recent neuroimaging experiments (McClure et al., 2004, 2007; Tanaka et al., 2004; see Fig. 4B and C). In one of these studies (McClure et al., 2004), human subjects chose between two delayed monetary rewards, where the larger reward was always more delayed than the smaller reward, and the delays ranged from immediate delivery to several weeks delay. When one of the choice options comprised a reward delivered immediately, or with a minimal delay, limbic brain areas, including ventral striatum, medial OFC and mPFC, were stronger activated than when both rewards were in the distant future. Moreover, lateral prefrontal and parietal areas were activated in all choice situations, independent of the delay, but when the subjects preferred the later reward over the more immediate, the activation magnitude in lateral prefrontal cortex was elevated. This result allowed to dissociate the contributions of two separate, distinct processes believed to correspond to the β- and δ-systems: limbic systems were selectively activated when the choices produced an immediate or short-term outcome, and therefore qualified as a potential candidate for the impulsive β-system. Lateral prefrontal and parietal areas were non-selectively activated, but had a stronger activation level when the subjects made a far-sighted decision, suggesting it could be a correlate of the long-run δ-system. Similar results have been found by Tanaka et al. (2004).

It is unclear how the different choice systems are supposed to influence a decision. The computational outcome of both systems may be combined into a single integrated utility function which could then be used to guide the decision. This is consistent with the evidence discussed above that the discounted value is represented as a single currency (Montague and Berns, 2002), for example in OFC (Roesch and Olson, 2005b) or NCL (Kalenscher et al., 2005b). Alternatively, a race-model is also conceivable in which all systems independently exert an influence on a decision network generating the categorical decision without being integrated into a single utility representation. According to this view, both systems compete with each other, and following a winner-takes-all principle, the stronger of both systems dominates the weaker system, and determines the decision. Evidence for such winner-takes-all choice mechanisms have been reported in perceptual and probabilistic decision-making (Kim and Shadlen, 1999; Yang and Shadlen, 2007).

Much of what we know about the neuroscience of intertemporal decisions stems from non-human animal research. To the authors' knowledge, there is no direct evidence from the animal literature to support this multiple-process hypothesis, although the above cited study by Roesch et al. (2006) may be interpreted in favour of this theory because of the differences in the sensitivity of the distributed cell populations to delay ('short-term system') and reward amount ('economic planning system').

In general, animals are substantially more impulsive than humans (Ainslie, 1974; Tobin and Logue, 1994; Green et al., 1994, 1996, 1997; Stevens et al., 2004). This may point towards a complete absence of a far-sighted choice system, implying that the multi-system hypothesis only inadequately captures animal behaviour. Alternatively, this disparity in human and animal behaviour can also be attributed to systematic differences in task configurations and requirements. Differences in experimental settings are likely to affect the way future rewards are discounted, as evidenced by a recent study that showed that the typically observed discrepancy in impulsivity between humans and apes disappears when tested with equal task parameters (Rosati et al., 2007). Typically, there is a huge difference in the scale of the delay spans used in human and animal research. The delays in human research are usually in the range of weeks to years or decades, whereas animal studies normally employ delays in the seconds to minutes range (cf., Clayton and Krebs, 1995, though, for very far-sighted behaviour in food-caching birds). Because of this, human subjects almost never experience the reward delivery during the course of the experiment. Instead, they are mostly instructed to imagine the delays. Animals in turn learn the delay lengths through experience, hence through repeated exposure to the delay and its end, as signalled by reward delivery. Imagining and experiencing delay lengths and rewards may have substantially different impact on reward valuation and discounting, for example, by triggering different levels of impatience. Moreover, whereas human subjects are usually rewarded with a strong, but abstract secondary reinforcer, money, animals receive a primary, appetitive reinforcer, usually food or liquid. Primary and secondary reinforcements invoke different psychological mechanisms, and recruit at least partially different neural networks (Bassareo and DiChiara, 1999; Parkinson et al., 1999; Grimm and See, 2000; Gottfried et al., 2002a, b; Estle et al., 2007).

To address whether the hypothesis of multiple systems in the brain also holds when primary rewards are involved and the delays are within the minutes range and actually experienced by the subjects, McClure et al. (2007) replicated their earlier study with human subjects, this time using fluid rewards and delays in the range of minutes, not weeks or months. They found a very similar activation pattern as in their 2004 study, and concluded that the same multiple-system account that they used to interpret their earlier results also applies when primary rewards are involved at much shorter time delays.

4.9. Challenging the multiple systems hypothesis

McClure et al. (2007) used a complex experimental design, and this complexity has to be considered when interpreting the results. In each trial, subjects decided between a large, delayed or a small, short-term fluid reward, and then received one or multiple fluid squirts, depending on earlier decisions. In many cases, the chosen reward in a given trial was only delivered several (up to 60) trials and choices later, and the total amount of fluid squirts received in a given trial was a composition of rewards that could stem from the immediately preceding trial, but also from several trials in the past. This design made it difficult for the subjects to associate outcomes with choices,

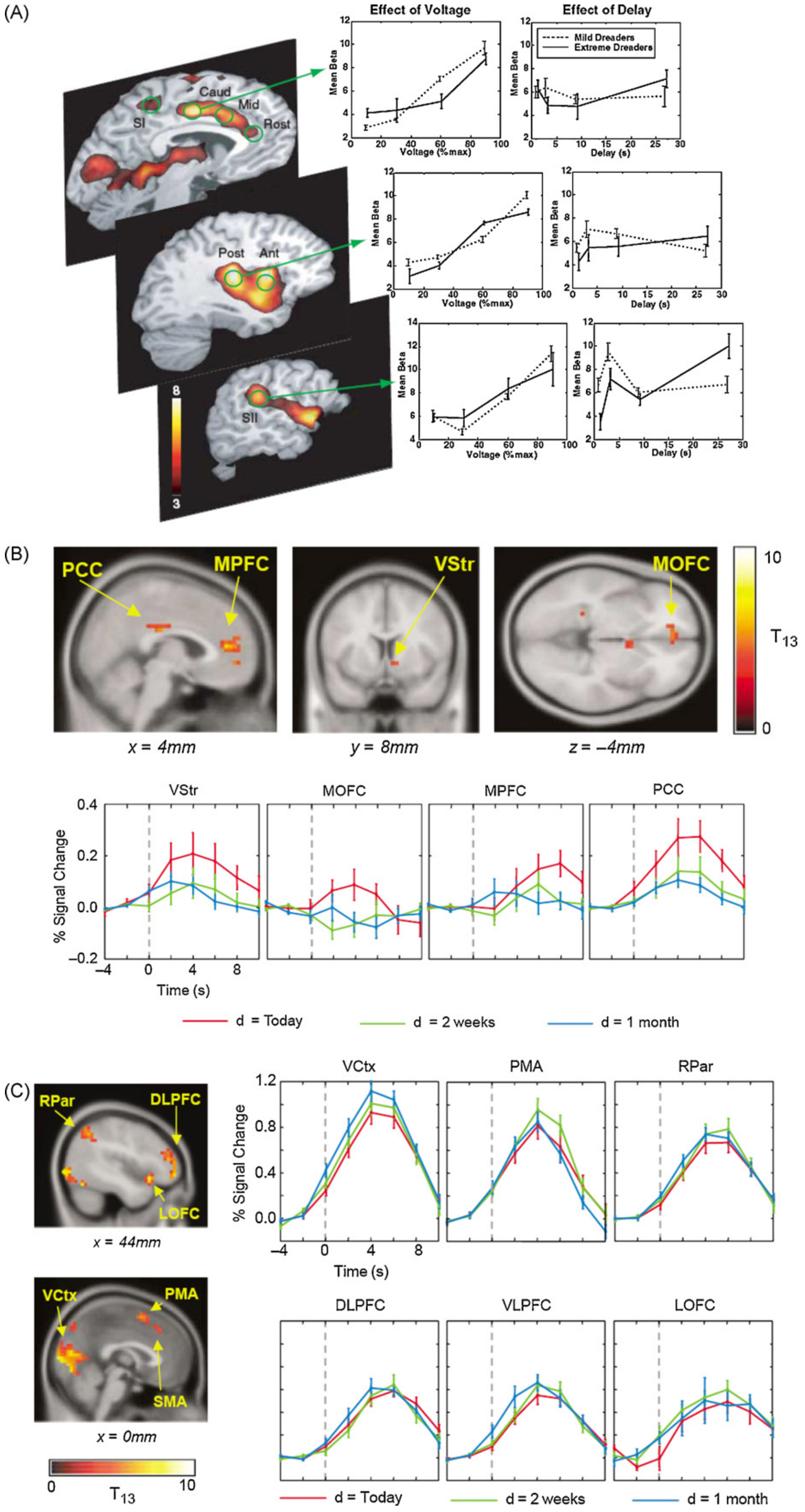

Fig. 4. Blood-oxygen-level dependent (BOLD) activation of different brain regions during intertemporal decision-making in the domain of aversive events (A, Berns et al., 2006), or in the domain of gains (B and C, McClure et al., 2004). (A) Effect of voltage and delay on BOLD responses in structures of the pain matrix, including primary and secondary somatosensory cortex (SI and SII), caudal, middle and rostral anterior cingulate cortex (Caud, Mid and Rost), and anterior and posterior insular cortex (Ant and Post). Although BOLD responses discriminated significantly between levels of shock intensity, there was no difference between high and mild

and they certainly did not experience the delay lengths in the same way as the animals did in the above-mentioned studies. Because the authors were not interested in associative learning, the difficulty in linking choices with outcomes does not necessarily invalidate their interpretation. But, it was inherent to the design that subjects often received fluid squirts immediately after their choice that originated from several choices in the past, even when they opted for delayed rewards. It is, hence, uncertain in how far the subjects were anxious to speed up the delivery of a reward if they received an immediate reward anyway, and it is unclear to what extent the subjects' decisions were actually guided by impatience, or other choice mechanisms with short time horizons, in particular towards the end of the experiment. Because short-run mechanisms are, however, essential for the interpretation of the neural results within the β–δ-framework, we feel that it is at least a matter of debate whether this study provided evidence for the multiple-systems account of primary reward discounting.

A recent neuroimaging study by Glimcher et al. (2007) provided further and more general challenge of the hypothesis of multiple systems in the human brain. The authors reasoned that, if the choice-determining utility function indeed results from combining a very steeply discounting β-system with a less steeply decaying δ-system, then the curve of the actual utility function should be somewhere in-between the β- and δ-functions. The critical test, hence, would be to show that the slope of the discount function fitted to the BOLD signal of the limbic areas believed to process the β-system is steeper than the slope of the discount function derived from behavioural measurements. Using a series of decisions similar to the ones employed in the McClure et al. (2004) study, Glimcher et al. (2007) individually measured the behavioural indifference points for different rewards and delays, and obtained neural discounting parameters by fitting discount functions to the BOLD signals of the limbic areas believed to process the β-system.

As expected, the best fitting functions to both the behavioural and neural data were hyperbolic, and not exponential. However, the slopes of the neurally measured functions on the single-subject level were not steeper than the slopes of the behaviourally estimated functions, even when immediate rewards were involved. Instead, both functions corresponded surprisingly well. This suggests that the BOLD-levels in limbic areas were linked to the actual behaviour, and not to the over-impulsive short-range system, as hypothesised by McClure et al. (2004, 2007).

In addition, Glimcher et al. (2007) reasoned that a short-term β-system should only be activated when the choice set includes an immediate or short-term gain, but not if all options in the set were in the relatively far future. In line with previous studies, the limbic regions believed to constitute the β-system were activated when subjects chose the temporally more proximal option, but this also held true when both outcomes were in the relatively far future, i.e., when no instant gains were involved. Hence, incompatible with the β-system interpretation of limbic activation, activity in those structures seemed to be more related to choices of 'as soon as possible' options than immediate options.

Taken together, these results challenge the view that the components of the multiple-systems theory can be mapped onto separate systems in the brain. Because of this, and the lack of evidence in animal research, we feel that the generality of the multiple-system models in neuroscience remains questionable, although we acknowledge that the ongoing debate could well be decided in their favour, at least in humans.

In the next section, we will present an alternative, biologically plausible model to account for temporal inconsistencies in choice. The model is based on the learning literature and the implementation of learning rules in the brain. In the model, time preference is treated as a unitary construct. This means that we assume that reward utility is discounted hyperbolically, and hyperbolic discounting holds even without necessitating the influence of additional cognitive processes, such as utility from anticipation, impatience, reluctance of delaying gratification, or other multiple systems.

We justify treating time preference as a unitary construct by referring to Occam's razor. Because of the uncertainty about a neural substrate of the multiple-system hypothesis, we feel that neuroeconomic phenomena can be better accounted for by simple learning rules than by assuming complex, sophisticated reckoning, unless compelling evidence is provided to the contrary (cf. Staddon, 2001; Barraclough et al., 2004). By no means, however, do we attempt to rule out in principle the possibility that time preference is a composite of several basic constituent processes, as impatience, delay of gratification and similar motives undoubtedly affect choice. We simply maintain that hyperbolic discounting can be achieved even without requiring multiple processes. We see our model as well as all other computational models discussed in the next section as first attempts to approach the issue of time preference. They can be easily extended by adding extra computational processes in future research.

Most of the learning literature on which the model is based on originates from animal research. It is a relevant question whether animal research can inform the science on intertemporal choice behaviour at all, given all the considerations about likely species differences. We do think so. First, economic decisions in animals are interesting in themselves, and entire research areas are devoted to them, e.g., optimal foraging theory. Second, as indicated, much of what we know about the neural implementation of intertemporal decisions,

dreaders. However, with the exception of the right SII, delay length had a differential effect on BOLD responses of high and mild dreaders (see text for details). Reprinted from Berns et al. (2006). Copyright 2006 with friendly permission from AAAS. (B) Activated brain regions for choices in which money is available immediately. PCC: posterior cingulate cortex; MPFC: medial prefrontal cortex; VStr: ventral striatum; MOFC: medial orbitofrontal cortex. (C) Brain regions that were activated while making choices independent of delay. RPar: right intraparietal cortex; DLPFC: dorsolateral prefrontal cortex; VLPFC: ventrolateral prefrontal cortex; LOFC: right lateral orbitofrontal cortex; VCtx: visual cortex; PMA: premotor area; SMA: supplementary motor area. Reprinted from McClure et al. (2004). Copyright 2004 with friendly permission from AAAS.

reward processing and learning theory stems from animal research. It is therefore straightforward to build intertemporal choice theories on existing knowledge from the animal literature. Third, animal research avoids many of the confounds that are a frequent problem in human science, e.g., human subjects' choices may be biased by their hypotheses about the aim of the experiments, the use of strategies of self-regulation is more difficult to control in human than animal experiments, task instructions may induce framing effects, choices may be biased by the actual inflation rate whenever money is involved, etc. Hence, because the confounds contained in this list are less likely to influence animal than human behaviour, chances are higher that animal experiments target a less biased time preference. We are therefore convinced that animal science provides highly valuable contributions to understand intertemporal choice.

5. Towards a neurocomputational model of intertemporal choice behaviour

5.1. Delay discounting and neurocomputational modelling

In the previous sections, we discussed, among others, how neuroscience can inform economics and psychology, i.e., in how far behavioural models of intertemporal choice can be validated with neuroscientific methods. Here, we attempt to devise an alternative neuroscientific and biologically plausible model of intertemporal choice.

The key to understand most anomalies discussed in Section 3 is to understand hyperbolic discounting. We postulate that the question how hyperbolic discounting is neurally implemented can be best approached by reducing it to the problem of reward value updating in the brain: how is the temporally discounted reward value computed and represented in the first place, how is it updated when the waiting time to a reward increases, and, most essentially, how is the subjective value updated hyperbolically? These issues can be addressed with neural mechanisms accounting for simple behavioural phenomena, namely Pavlovian learning (Rescorla and Wagner, 1972). The adoption of learning models as a foundation for intertemporal choice behaviour is a straightforward and logical one, because it would be hard to consider the updating of reward value as a process fundamentally different from learning. By consequence, all models presented below assume that animals will be offered sufficient amounts of learning trials to update their values appropriately, regardless of whether they pertain to immediate or delayed rewards. In addition to flexibly updating subjective reward values, an animal must also be able to predict the values of the different response options in order to weigh the options, and come to a categorical decision. Hence a viable model must also account for how the reward value representations are invoked by external or internal events preceding the choice, such as conditioned stimuli, cues, thoughts, action and homeostatic variables. Once we have found a model that explains the computation, representation and hyperbolic updating of value representation, and its attribution to stimuli or events preceding choices, it can be extended and generalised to more complex situations, such as decisions involving different cue and action values (Dayan and Balleine, 2002).

Because of hyperbolic discounting, our and also some other considered models can explain preference reversals and the common difference and immediacy effects. Furthermore, because of the delay-sensitive valuation system, the choices predicted by the models will follow a short-term rule which results in the failure to maximise reward rate in intertemporal choice situations in which trial lengths are equal.

However, because there is little evidence that rewards recruit the same neural circuits and mechanisms as losses or aversive events, we will restrict this section to the discussion of the appetitive learning literature. This means that the deliberated models cannot necessarily be extrapolated to intertemporal choices in the domain of losses. Hence, the discussed models are restricted to anomalies in the domain of gains, and they cannot readily account for sign-effects, or the acceleration of aversive events.

5.2. Temporal difference learning

Previously, models of neural systems computing reward values based on learning experiences have been proposed in various forms (e.g., temporal difference learning, TD; Sutton and Barto, 1987; Sutton, 1988; Schultz et al., 1997), but these have not specifically focused on explaining the intertemporal choice patterns in animals and humans described above. In this section, we will therefore briefly review some general requirements and constraints on computing reward value in a time-dependent manner and then focus on specific neural models accounting for temporal discounting and preference reversal.

A central concept in modelling choice behaviour is that sensory cues, contexts and actions can acquire a certain time-dependent value due to the associations of these inputs with rewards or punishments received later in time. The general form of a value function can be expressed as

$$V_t = E[\gamma^0 r_t + \gamma^1 r_{t+1} + \gamma^2 r_{t+2} + \cdots] \quad (8)$$

where V_t denotes the value that an animal associates with the cues and context present at time t, E the expectation (or prediction) of all rewards expected at the current time t and in the future, as represented by the terms in the square brackets (cf. Sutton and Barto, 1987; Schultz et al., 1997; however, the definition of value function V_t used here is different from that in Schultz et al., 1997). In this equation, rewards r are predicted to occur at discrete moments in time (t, $t + 1$, etc.) and are temporally discounted according to the γ factors, which decrease in magnitude as the temporal distance to t increases (Sutton and Barto, 1987; Sutton, 1988). A neural system facing the task of learning to make choices should be able to utilize experiences involving temporal associations between neutral environmental events and valuable outcomes in order to generate an accurate estimate of how much reinforcer will become available in the future, and when in the future this will occur. If the subject is confronted with multiple response options in the situation at hand, reward estimates can be computed for each relevant environmental event or context. Reward quality is not considered separately from quantity in the current model, nor is the impact of negative reinforcers on learning explicitly considered, although both of these aspects could be included in more elaborate versions. Eq. (8) can be changed to the following equation:

$$V_{t-1} = r_t + \gamma V_t \tag{9}$$

which states that the value function V at time $t-1$ is computed as the reward occurring directly at time t (r_t) added to the (discounted) sum of all future rewards expected to occur later than t (Sutton and Barto, 1987). The use of an $E[\ldots]$-term can be omitted from the equation since the reward r_t occurs at moment t of evaluation, and V is expectational by definition. Eq. (9) will only hold in the case of perfect reward prediction, i.e., when the animal's value function at time $t-1$ exactly matches the reward arriving one time step later and the sum of all expected future rewards. In cases of mismatch between actual outcome and prediction an error term δ_t can be computed by bringing V_{t-1} to the right-hand side of the equation:

$$\delta_t = r_t + \gamma V_t - V_{t-1} \tag{10}$$

In TD learning (Sutton and Barto, 1987; Sutton, 1988), the error in reward prediction δ_t will be larger than zero at time t if the reward occurring at t is larger than predicted by the term ($\gamma V_t - V_{t-1}$), such as when a naive animal unexpectedly encounters a reward whilst not having learned yet that it is regularly preceded by a conditioned stimulus (CS). Similarly, δ_t will be negative if an expected reward does not occur. Moreover, error terms can be computed even before the neural system has perceived an actual reward in a given trial. This happens when there is a moment-to-moment change in value function so that ($\gamma V_t - V_{t-1}$) is different from zero. Thus, not only does δ_t provide a prediction error signal upon receiving a reward, it can also deliver a 'surrogate prediction error' (Schultz et al., 1997) when, for instance, a relevant CS or action occurs.

The TD-learning model has a remarkable formal resemblance to the firing properties of dopaminergic neurons in the monkey ventral mesencephalon (cf. Schultz et al., 1997; Hollerman and Schultz, 1998) and several neural architectures have been proposed for representing and computing the quantities of Eq. (10) (Schultz, 1998; Beiser et al., 1997; Montague et al., 1996; Montague and Berns, 2002; Daw and Doya, 2006). Generally, we hypothesize value representations V_t to be generated by higher-order associative brain areas that process sensory input pertaining to the identity of conditioned stimuli and convert this information into reward-predicting neural signals. Such areas will be referred to as 'sensory-evaluative'. In a TD-learning scheme, their representational output is subjected to a discounted temporal-differentiation operation to compute ($\gamma V_t - V_{t-1}$) and emitted to one or more target areas; the differentiation may take place either within the sensory-evaluative area or in the target area. When this differentiated output signal, reflecting the moment-to-moment change in reward prediction, is then summed in the target area with an input reporting the actual reward value r_t, the discrepancy (error) between actual and predicted reward can be computed (Eq. (10); Fig. 5). Once this error term is available, it can be broadcast into the sensory-evaluative brain regions to guide further learning of CS-reward associations and it can be emitted to brain structures for decision-making and response execution.

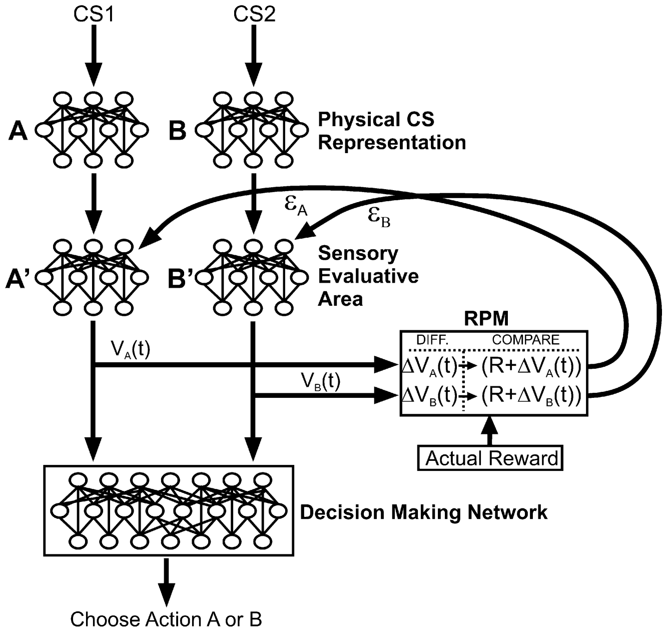

Fig. 5. Schematic diagram of a general layout for a neural network architecture performing intertemporal choice behaviour based on temporal difference learning. The scheme departs from the situation where an organism faces two environmental cues (CS1 and CS2) and must choose an action A or B to react to these two cues. Following a neural stage where the physical features of CS1 and CS2 are represented by the neural ensembles A and B in sensory cortical areas, the CS1 and CS2 information is transmitted to two separate neural ensembles A′ and B′, which each generate a value representation that varies as a function of time ($V_A(t)$ and $V_B(t)$). These ensembles convert sensory input into an evaluative signal and are hence termed 'sensory-evaluative'. The efferent fibers from ensembles A′ and B′ are routed both towards a decision-making network which will favour the action associated with the highest current value ($V_A(t)$ and $V_B(t)$), and towards a reinforcement-processing module (RPM). The RPM first computes the first differentials $\Delta V_A(t)$ and $\Delta V_B(t)$ as a function of time, as indicated by the operation "DIFF". At a neural level, this operation corresponds to converting the tonic outputs from the ensembles $V_A(t)$ and $V_B(t)$ into phasic signals. Next, the RPM performs a comparative operation ("COMPARE") by summing the $\Delta V_A(t)$ and $\Delta V_B(t)$ signals with the signal representing the actual reward at time t. The resulting error signals ε_A and ε_B are next fed back to the sensory-evaluative ensembles A′ and B′, and are necessary to enable the network to learn to predict future reinforcement and to improve decision-making. The ensembles A′ and B′ may be located in the same brain region (indicated by the dotted ellipse) or in different brain regions. Note that the implementation of this general scheme is not based on a particular transmitter system mediating the error feedback, such as dopamine or glutamate.

Due to its presumed function in guiding learning, δ_t is considered a (scalar) 'teaching signal' and a computationally efficient learning rule capturing this role in a multi-layer neural network is given by:

$$\Delta w_{ij,t} = c\delta_t a_{i,t} a_{j,t} \tag{11}$$

where $\Delta w_{ij,t}$ denotes the change in the synapse from presynaptic neuron j to postsynaptic neuron i in a single- or multi-layered network at time t, c is a learning rate constant and $a_{i,t}$ and $a_{j,t}$ represent the (firing) activity of the post- and presynaptic neuron at time t, respectively (cf. Sutton and Barto, 1987; Schultz et al., 1997; Schultz, 1998; Montague et al., 1993; Pennartz, 1996; note, however, that no postsynaptic term $a_{i,t}$ was included in the original TD-model of Sutton and Barto, 1987). This rule may be applied to multi-layered networks in sensory-evaluative as well as executive brain areas.

Since its original development in the field of robotics and control theory, TD-learning has gained significant merit, not only in neuroscience but also due to its broad applicability to cognitive problems both in psychology and artificial intelligence (e.g., learning to play backgammon with TD-gammon, Tesauro, 1994). Especially noteworthy in the present context is the ability of TD-models to mimic a number of phenomena encountered in conditioning paradigms, such as blocking, inhibitory backwards conditioning, stronger conditioning induced by forward CS–US pairings rather than simultaneous CS–US presentation, second-order conditioning and the formation of novel associative chains (Sutton and Barto, 1981, 1987, 1990; Malaka and Hammer, 1996; Suri, 2001; Seymour et al., 2004).

At this point, an important distinction should be made between the general computational principles pointed out above, such as the concept of a value function and an error in reward prediction on the one hand, and on the other hand the specific implementation of these concepts in a particular model and learning rule (i.e., TD-learning) and their attribution to particular neural systems (i.e., the midbrain dopamine system with its set of afferent and efferent connections and structures). Because it is, for instance, empirically uncertain whether dopamine will affect synaptic plasticity bidirectionally as predicted by Eq. (11), or which forms of reward-dependent learning are driven by dopamine, it is important to realize that a learning system for reward prediction may implement the general concepts by other means, using different algorithmic rules and different architectures (cf. Pennartz, 1996). To illustrate this, a neural architecture for reward prediction learning has been constructed that avoids a ‘triadic’ learning rule (Eq. (11)) and instead uses Hebbian modification of glutamatergic synapses with adaptive plasticity thresholds (the Hebbian synapses with adaptive thresholds-, or HSAT-model; Pennartz, 1997). This model employs glutamatergic projection neurons to convey actual reward values as well as reward-predictive signals, in agreement with the glutamatergic nature of pyramidal neurons throughout the neocortex and additional structures implied in evaluation, e.g., the BLA. As in a dopaminergic instantiation of reinforcement learning, cortical and amygdaloid neurons are presumed capable of representing reward values upon arrival of predictive sensory cues, as reviewed above, and local microcircuitry either within these sensory-evaluative areas or in target areas is assumed to carry out a differentiating operation to compute moment-to-moment fluctuations in reward prediction. Hence, besides the sensory-evaluative modules in the general scheme of Fig. 5, the error-computing module may be considered dopaminergic or glutamatergic and cortical or quasi-cortical (BLA) in nature.

5.3. *Neural implementations of a discounting function: TD model*

Let us next consider how a hyperbolically decaying discount function, as an exemplar accounting for intertemporal preference reversals, may be implemented in an efficient and neurobiologically realistic fashion. It is straightforward to compute the γ-factors in Eq. (8) according to a hyperbolic decay:

$$\gamma^t = \frac{\gamma^0}{1+kt} \tag{12}$$

where γ^t is the discount factor at time step t relative to an initial moment $t = 0$, k a positive constant regulating the rate of decay and γ^0 the discount factor at time $t = 0$ (see also Eq. (6)). Other decay functions can be used that will result in a steep and asymmetric decay and preference reversal (i.e., cross-over in two value functions for choice options with rewards at different delays).

However, it is less straightforward to point to a prime candidate neural mechanism imposing such a function on the weighing of temporally proximal *versus* distal events. Below we will evaluate four possible mechanisms bearing relevance both for a ‘dopaminergic’ and ‘glutamatergic’ (or still other) neural implementation.

Let us first consider whether the TD-model may offer a plausible neural implementation for appropriately weighing temporally proximal *versus* distal events. Throughout each trial the TD-learning rule (Eq. (11)) will be continuously applied within the network, even if no predictive CS or reward is presented; weight changes will be induced whenever a positive or negative prediction error signal coincides with pre- and postsynaptic activity and this can already occur when γV_t differs from V_{t-1} (Eq. (10)). In the learning rule, the discount rate γ is only applied to the most recent value V_t but not to the previous value V_{t-1}. By itself, such a discounting mechanism may be neurally implemented by computing a phasic signal assigning a slightly higher amplitude (if, e.g., $\gamma \sim 0.95$) to the previous value V_{t-1} than the current value V_t, which is possible using, e.g., feed-forward inhibition. There is, however, a more fundamental problem in seeking a straightforward neural implementation. When a CS occurs, the model assumes that this event will be represented by many different synaptic weights, one for each time point following CS presentation. Thus, an expanded ‘time register’ would be set up for each CS in a time-discrete manner, such that each time point of each CS is assigned its own, specific reward-predicting weight. Under this scheme of a ‘serial-compound stimulus’ (Sutton and Barto, 1987), a CS is represented as a long vector of signals, each of which represents the cue at a different time interval into the future. Given the evidence in favour of temporally graded neural representations for coding of motivational value and time intervals (Watanabe, 1996; Tremblay and Schultz, 1999; Komura et al., 2001; Montague and Berns, 2002; Durstewitz, 2003; Kalenscher et al., 2005a, b, 2006b; see further below), such a time-discrete register cannot be considered neurobiologically realistic. For instance, neurons in sensory or higher-associational areas do not respond to sensory stimuli in a block-like, time-discrete manner, at the same time representing a multitude of time delays for one and the same sensory stimulus. Representing each and every stimulus by a long vector of signals across a range of time intervals would pose enormous demands on neural coding and storage capacities, which are large but not endless. This problem may be overcome if more biologically plausible stimulus representations will be identified to provide time-specific CS-value signalling required for TD-learning. Thus, the problem in implementing an appropriate discount function by TD-learning may be more related to the way stimuli are represented than to a fundamental limitation of the rules that govern the dynamics and learning in TD-models.

5.4. Neural implementations: climbing-firing profile

Instead of applying a TD-model, it is worth considering whether ramping or 'climbing' firing profiles of neurons in limbic and striatal structures may provide a neural mechanism for implementing an appropriate temporal discount function. As outlined in Section 4.2, many reward-related structures, including OFC, cingulate cortex, amygdala and striatum, contain large subsets of neurons displaying upward 'ramps' in firing rate, with their onset usually shortly after a CS and their peak rate just in advance of the reward or motor action resulting in reward (Fig. 6B; Niki and Watanabe, 1979; Schultz et al., 1992; Watanabe, 1996; Quintana and Fuster, 1999; Rainer et al., 1999; Tremblay and Schultz, 1999; Brody et al., 2003; Kalenscher et al., 2006b). Contrary to a decaying eligibility trace which partly retains cue information as time elapses (see Section 5.5), an upward ramp in firing rate may be thought of as a 'forward' trace specifying the temporal proximity to a reinforcer. In their predictor-valuation model that aims to explain how reward predictors acquire value, taking into account both the uncertainty of a future reward and the risk of allocating computational resources to a predictor of the reward, Montague and Berns (2002) proposed this escalating increase in firing rate as a neural mechanism for a valuation function.

However, if a climbing-firing profile represents a temporally discounted value of some reward predictor that has the capacity to guide decision-making, a problem with this mechanism would be that the mean firing rate is usually at or close to zero at the moment of CS presentation. Animals are clearly able to choose between two CSs as they are presented, and thus their neural systems must be able to quickly retrieve their reward values from memory and represent them actively. Thus, ramp functions emerge too slowly in time to enable decisions at the time of predictor presentations. A second problem for ramp functions is that in many climbing-firing profiles the mean firing rate usually drops towards zero shortly after or even before reward delivery. Regardless of the precise learning rule used, it is essential that neural activity signalling actual reward coincides in time with active neural representations of reward-predictive value. Eq. (11) illustrates this principle by computing weight changes from the product of the error in reward prediction, a term including actual reward (Eq. (10)), and of pre- and postsynaptic activity, involved in value representation. If the information on actual reward only arrives after the value representation has collapsed, synaptic modifications in the value-representing network cannot be guided by real-world reinforcement.

Instead of a function of climbing-firing profiles in temporally discounted value representation, we adopt the view that these profiles serve a more general function, viz. the coding of time intervals in between behaviourally important events such as CSs and outcomes. The idea that a neuron's instantaneous firing rate can be used to code the time elapsed since stimulus presentation and the time remaining until trial outcome has been elaborated in neural integrator models for prediction of interval times (Durstewitz, 2003). As explained above, one of the predictions from these models holds that the rising rate of climbing functions may be flexibly adjusted, which has been confirmed by experimental observation of such adjustments when CS-reward timing intervals change during learning (Komura et al., 2001; Kalenscher et al., 2006b).

5.5. Neural implementations: eligibility trace

A second mechanism for weighing temporally proximal *versus* distal events is illustrated in Fig. 6A and concerns the possibility that information about neutral sensory events is temporally stored in a neural structure by inducing an 'eligibility trace' (Sutton and Barto, 1987). First we note that the initial storage of CS information before arrival of a reinforcement after a time delay is a general requirement for all conditioning processes that span considerable time intervals, regardless of its precise form, dynamics or mechanism. Second, the neural substrate of such a trace can be envisaged as, for instance, an elevation of the intracellular pre- or postsynaptic concentration of a second messenger (e.g., cAMP) or the degree to which a synaptic protein, relevant for long-term plasticity that may ensue in a subsequent phase, is phosphorylated (e.g., a subunit of the NMDA receptor, or Ca^{2+}/calmodulin-dependent protein kinase II, CaM kinase II; Rosenblum et al., 1997; Sweatt, 2001; Lisman et al., 2002; Nakazawa et al., 2006; Gervasi et al., 2007). When a CS occurs at time $t = 0$ (Fig. 6A), the trace is first presumed to peak rapidly and then, following CS offset, decay back towards baseline level. The strength of the trace represents the time-dependent activity value $a_{j,t}$ (or $a_{i,t}$) pertaining to a CS or context (cf. Eq. (11)), and its decay implements discounting. If a rewarding event subsequently occurs at a time $t = \Delta$ when the strength of the trace is still significantly elevated above baseline, the trace-activated synapses will undergo a long-term modification according to, e.g., the HSAT learning rule for reinforcement learning. Such discounting by trace decay could be broadly applicable to learning models, although in the TD-model a discount mechanism was already incorporated in the computation of errors (Eq. (10)).

At present, an eligibility-trace mechanism for implementing temporal discounting seems a neurobiologically plausible model, although admittedly there is little experimental

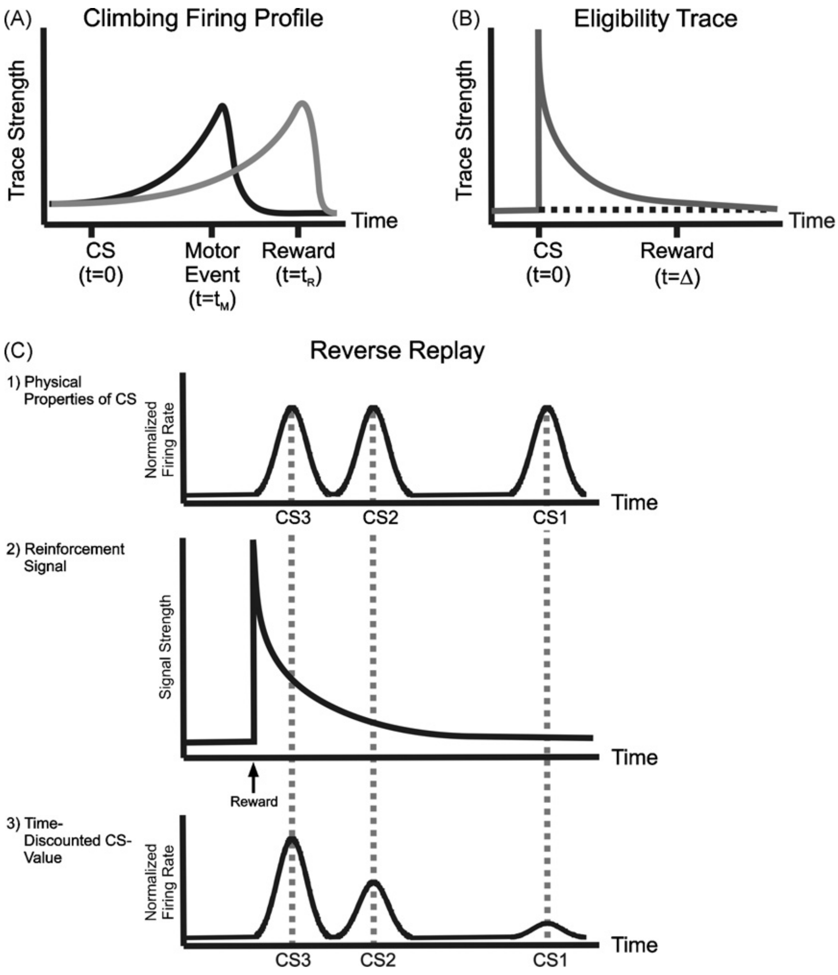

Fig. 6. Three possible neural mechanisms for weighing the relative value of temporally distal vs. proximal events. Such mechanisms are postulated to be part of models of intertemporal choice behaviour because of the requirement to assign a lower motivational value to delayed as compared to temporally proximal reward. A requirement for these mechanisms is that they can generate a time-dependent asymmetric, hyperbolic decay in event value in order to account for reversal of choice preference. (A) Climbing firing profiles, such as encountered in limbic, cortical and striatal structures, are characterized by a slow increment in firing activity that is initialized around the moment of CS occurrence ($t = 0$) and is terminated around the time the animal undertakes an action ($t = t_M$) or a reward is delivered ($t = t_R$). The time-dependent firing rate of the neuron can be taken to represent the strength of a trace coding the relative value assigned to events occurring in advance of the reward. The value is constantly increasing due to the continuously decreasing residual waiting time, and hence increasing temporal proximity to a predicted reward while waiting for it. The black and grey curves illustrate climbing firing profiles with end points locked to the motor event (black curve) or reward (grey curve), respectively. (B) An 'eligibility trace' may correspond to a biochemical signal in the pre- or postsynaptic elements of a set of neurons that rapidly reaches its peak strength once the reward-predicting event (CS, at time $t = 0$) occurs and subsequently decays steeply towards baseline level. The value assigned to CS will be determined by the trace strength that remains at the time Δ the reward occurs. (C) A 'reverse replay' scenario is conveniently explained by distinguishing three levels of event processing: at a first stage (C1), the physical features of relevant events (CS1–CS3) are represented in sensory cell ensembles (see also Fig. 7). The graph plots three events across time, with CS1 being the earliest event and CS3 the one most proximal to reward. The order of the CS occurrences is reversed since the model assumes that these events are replayed backwards in time. The peak firing rates have been normalized across all predictive events. Hence, CS3 occurs earlier on the time axis with respect to reward occurrence than CS2 and CS1. Graph C2 represents the strength of a reinforcement signal as a function of time, with the reward occurring as indicated by the arrow. Note that all panels (C1–C3) are temporally aligned with one another, so that the reward signal can be seen to be initiated shortly before the replay of the CS3-representation. In C2, the decay of the reinforcement signal assumes an asymmetric, hyperbolic shape. C3: if the time-dependent values of the physical CS-representations (C1) are essentially multiplied with the strength of the reinforcement signal applicable to the corresponding moments in time (C2), an appropriate weighing of the various CS events can be obtained. An essentially Hebbian learning rule may suffice to perform this multiplicative operation (Eq. (13), see text).

evidence that clearly argues in favour of hyperbolically decaying neurochemical traces. Although steeply decaying transients in pre- or postsynaptic [Ca^{2+}] and cAMP or other second-messenger traces have been reported (Hempel et al., 1996; Scheuss et al., 2006; Majewska et al., 2000), their exact decay dynamics at the relevant sites of synaptic modification remain to be resolved and it is questionable whether the longevity of such traces is sufficient to span time intervals that are relevant to the time scale of delayed reward effects on behaviour. Phosphorylation of CS-activated pathways may provide a more appropriate temporal dynamics, but little is known about the exact decay dynamics at behaviourally relevant time scales. A potential problem with this model is that synaptic weight changes tend to be effectuated by short-latency interactions between pre- and postsynaptic elements, ranging in the order of 20–100 ms (Levy and Steward, 1983; Markram et al., 1997; Bi and Poo, 1998). This disadvantage would not apply if a neural mechanism could be identified that is able to pack relevant CS- and reward-related information tightly together in time and uses well-known principles of Hebbian synaptic plasticity applicable to short time frames. Another disadvantage of a trace mechanism is that the storage of relevant one-shot associative experiences (e.g., pairing taste to nausea in conditioned taste aversion) depends on a singular mnemonic event without the benefit of rehearsal. Although a potential mechanism with this capacity is indicated below, it should be stressed that an eligibility trace-model remains a possible and perhaps plausible mechanism for temporal discounting as well.

5.6. Neural implementations: reverse replay

A fourth mechanism (Fig. 6C) for appropriate temporal discounting refers to a recent suggestion that, following an initial learning experience during bidirectional track-running for food reward, firing patterns characteristic for this experience can be replayed off-line and in reverse temporal order (Foster and Wilson, 2006). Traditionally, replay is viewed as the spontaneous recurrence, during periods of rest or sleep (“off-line” states), of neural activity patterns that are characteristic of behavioural experiences preceding the rest-sleep period. Because of the high degree of specificity of pattern replay and the preservation of temporal-sequence information displayed by ensembles during behavioural experiences, replay is considered an important candidate mechanism for mediating memory consolidation (Skaggs and McNaughton, 1996; Nadasdy et al., 1999; Lee and Wilson, 2002*)*. In contrast to forward replay occurring particularly during slow-wave sleep, Foster and Wilson (2006) discovered a temporally reversed replay in the awake state, particularly during ripple-sharp wave complexes, flanking or overlapping reward consumptions intermittently occurring during track running. These results have fuelled the hypothesis that reward-dependent memory consolidation may rely on a time-compressed reiteration of place- and event-related information (Pennartz et al., 2002), accompanied by a time-weighed reinforcement signal such as a transient, steeply decaying release of a neurotransmitter, e.g., dopamine (Berridge and Robinson, 1998, 2003; Schultz, 1997, 2002, 2004; Schultz et al., 1997; Tobler et al., 2005) or glutamate (Pennartz, 1997; Pennartz et al., 2000; Lavin et al., 2005; Seamans and Yang, 2004). Further support for this hypothesis has come from evidence for reactivation in the ventral striatum, which has been observed selectively in neuronal subgroups that show modulation of their firing rate in close association with hippocampal ripple-sharp waves (Pennartz et al., 2004). If it is assumed that the reverse replay commences with a peak in reinforcement signal (Fig. 6C) as well as with the neural activity representing the event most recently experienced during preceding active behaviour, and furthermore, that the replayed event representation becomes associated with the value conveyed by the reinforcement signal, then it can be seen how events temporally more and more remote from the reinforcement (but now replayed backwards in time) are associated with a possibly hyperbolically decreasing reinforcement signal. Thus, the discount function is imposed by the forward-temporal decay of the reinforcement signal, coupled to a backward-temporal, compressed replay of context- and event information, so that temporally remote events are associated with a lower value than temporally proximal events. Although modelling the kinetics of such a reinforcement signal is underconstrained, a steep and asymmetric decay function resembling a hyperbolic decrease is not implausible from a neurobiological viewpoint. For instance, the reinforcement signal may be expressed by the release of a transmitter, e.g., dopamine, followed by its diffusion to receptors, active re-uptake and/or breakdown in the extracellular space. In the case of dopamine, transient peaks in $[DA]_e$ in striatum as gauged with fast-scan voltammetry suggest steep, possibly hyperbolic decay kinetics (Heien et al., 2004), which is confirmed by computational modelling of spatiotemporal patterns of diffusion and re-uptake processes of dopamine in the striatum (Cragg and Rice, 2004). Similar diffusion, re-uptake and turnover schemes are likely to hold for other types of neuromodulators, some of which are likely to cover different time scales than dopamine. A potential problem in this type of discount mechanism may be the limited time span of the behavioural sequence that can be replayed in reverse, despite the fact that temporal compression of the sequence representation likely occurs (Kudrimoti et al., 1999; Nadasdy et al., 1999; Lee and Wilson, 2002; Foster and Wilson, 2006). Another point of concern is that reverse replay has thus far only been demonstrated for the hippocampus, while our hypothesis extends the phenomenon to other brain areas. While it indeed needs to be examined whether reverse replay can be generalized to other brain structures, we do note that replay in extrahippocampal areas has at least been shown to be temporally coordinated with the hippocampus (Qin et al., 1997; Pennartz et al., 2004; Ji and Wilson, 2007; Lansink et al., 2007).

In addition to dopamine, glutamate has also been suggested as carrying reinforcement information, as indicated above (Pennartz, 1997; Pennartz et al., 2000). At first glance, implementation of a glutamatergic reinforcement signal appears unlikely under the scenario of reverse replay, since the diffusion and uptake kinetics for glutamate release range in the order of milliseconds (Keller et al., 1991; Hestrin, 1992; Silver et al., 1992). However, a longer lasting decay of reinforcement signal may be obtained in other ways than purely by transmitter diffusion, uptake or turnover. The time course of the signal can also be sculpted by the firing patterns of reinforcement-signalling neurons. In this context, it is striking to note that many single-cell firing-rate responses to reinforcers in structures such as the OFC, amygdala, cingulate and mPFC and striatum can be quite long-lasting, with their onset usually at or shortly after reinforcer delivery and with variable time courses lasting hundreds of milliseconds up to more than 10 s, outlasting the ingestion phase (Niki and Watanabe, 1979; Apicella et al., 1991; Shima and Tanji, 1998; Mulder et al., 2003; Roitman et al., 2005; Van Duuren et al., 2007). These post-reward responses could not be attributed to particular motor behaviours of the animal, such as licking. The rate and curvature of the decay in post-reward firing varies from neuron to neuron, and thus far it has not been investigated what type of decay is displayed by the overall population response of these neurons. If these prolonged reward-responsive neurons have a reinforcement-signalling capacity, and they are accompanied in time by event- and place-related firing patterns generated in advance of a relevant outcome, temporally discounted learning can be realized by reverse replay under this scenario. Thus, if a forward-decaying reinforcement signal, now constituted by the firing rate-profile of reward-responsive cells in neocortical and amygdaloid areas, would be reactivated together with a reversely replayed sequence of events (CS3, CS2 and CS1 in Fig. 6C), an appropriate temporal discounting of events may be achieved.

5.7. A model of intertemporal choices based on reverse replay and Hebbian learning

In conclusion, the first two mechanisms reviewed above – TD-learning with discrete time registers and climbing functions – have distinct disadvantages in terms of neurobiological plausibility or learning efficiency. The third mechanism, an eligibility trace holding CS-related information, may be considered plausible, but currently lacks empirical support and has disadvantages in terms of efficacy. A ‘reverse-replay’ scenario may at present be considered the most promising candidate mechanism, even though the phenomenon should be scrutinized much more deeply.

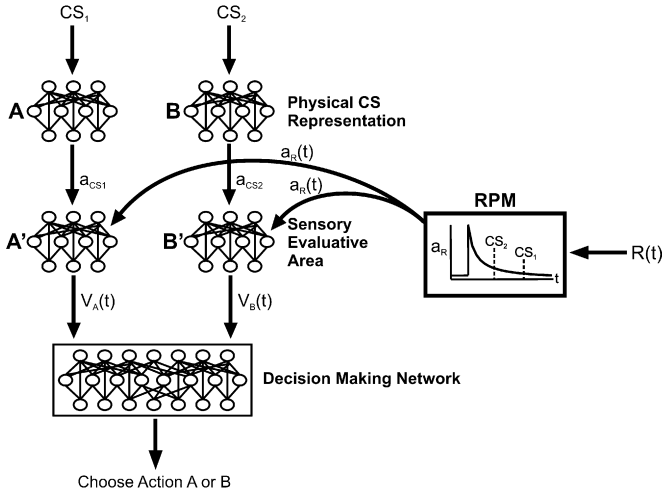

Fig. 7. Model for temporally discounted reward learning using reverse replay and Hebbian modification: network diagram implementing the reverse-replay operations laid out in Fig. 6C. As in Fig. 5, we only assume the processing of two stimuli for reasons of simplicity, CS1 and CS2, that are processed by two sensory ensembles A and B representing their physical features. In this example, CS1 occurs first in time, followed by CS2 and subsequently by the reward represented by $R(t)$. The outputs of A and B, a_{CS1} and a_{CS2}, are fed into the sensory-evaluative ensembles A′ and B′ which generate time-dependent value functions, $V_{\mathrm{A}}(t)$ and $V_{\mathrm{B}}(t)$, as their outputs. These value signals are used by the decision-making network to choose an action A or B, e.g., to approach CS1 and neglect CS2. Similar to the general scheme of Fig. 5, a reinforcement-processing module (RPM) transmits reinforcement signals to the evaluative ensembles A′ and B′ on the basis of its reward input $R(t)$, but different from Fig. 5, the value signals $V_{\mathrm{A}}(t)$ and $V_{\mathrm{B}}(t)$ are not fed into the RPM. In reverse-replay mode, the CS-related information (a_{CS1} and a_{CS2}) is regenerated in reverse temporal order, so that a_{CS2} is temporally coupled to a relatively strong reinforcement signal (a_{R}) and a_{CS1} to a relatively weak one, as illustrated in the graph inside RPM with hyperbolic curve. A Hebbian learning rule (Eq. (13)) is used to convert these temporal couplings into appropriate weight changes in the ensembles A′ and B′.

Fig. 7 presents an overall scheme capturing the main elements of the model as laid out above, and integrating these with additional elements for intertemporal decision-making. The scheme departs from the presentation of two conditioned stimuli (CS1 and CS2) coupled to different rewarding outcomes at different delays. The model only includes an evaluation of the reward parameters *amount* and *temporal proximity* to the CS, while other parameters such as reward probability are left out for simplicity. Above we already touched upon the as yet undecided question of whether reward amount and temporal proximity may be integrated into a common neural representation capturing both of these aspects (“common currency”), or whether these parameters may be separately represented in the activity of different ensembles. Acknowledging the remaining uncertainty on this issue and the possibility of different short- and long-run systems (McClure et al., 2004), Fig. 7 assumes an integrated representation. At the top end, two ensembles A and B involved in the sensory processing of CS1 and CS2 are shown; these ensembles represent the physical, non-motivational properties of the stimuli, whereas the lower groups of neurons represent sensory-evaluative ensemble A′ and B′ responsive to CS1 and CS2, respectively. A′ and B′ will process the CS1 and CS2 inputs from the overlying sensory areas to generate as output their reward-predicting value functions $V_{\mathrm{A}}(t)$ and $V_{\mathrm{B}}(t)$, which are propagated to a downstream decision network. In a first, straightforward implementation of the scheme (Fig. 7) the ensembles A′ and B′ are trained to generate appropriate values because their input synapses are modified to represent a running average of reward value. This training can occur because the postsynaptic neurons in A′ and B′ not only receive CS-information from the overlying sensory ensembles but also the reinforcement signal, emitted by a reinforcement-processing module depicted as the rightmost unit in Fig. 7, and assuming a steeply peaking, slowly decaying form (Fig. 6C). The following equation suffices as a basic learning rule:

$$\Delta w_{ij,t} = \alpha a_{r,t} a_{j,t} \tag{13}$$

where $w_{ij,t}$ is the weight of the connection from neuron j onto i at time t, α is a rate constant scaled between 0 and 1, $a_{r,t}$ the activity of the reinforcement signal emitted by the reinforcement processing module (RPM) and reaching the postsynaptic cell i at time t, and $a_{j,t}$ the activity of the CS-representing signal from the sensory ensemble A reaching postsynaptic cell i in A′, which can be normalized to 1 if the CS is occurring (otherwise $a_{j,t} = 0$). According to Eq. (13), weight changes in the sensory-evaluative area (Fig. 7) are directly dependent on the product of the neural activity representing the reinforcement signal and the CS. Despite the temporal separation between the CS and reinforcement, the two signals can be associated because the reinforcement signal decays slowly over time, as shown in Fig. 6C (middle panel). Thus, at the moment that a particular CS (e.g., CS2 in Fig. 6C) is reversely replayed, weight updating will start to occur for that particular CS proportional to the strength of the corresponding CS-representation and proportional to the strength of the reinforcement signal at the same point in time. To enable asymptotic learning in the network, the reinforcement signal $a_{r,t}$ is representing the unexpected (rather than absolute) amount of reward, or equivalently, the difference between the actual amount of reward minus the mean amount of previously received reward given a CS (cf. Pennartz, 1997).

If the sensory afferents as well as projections from RPM are glutamatergic, this learning conforms to a Hebbian scheme on account of the following. First, the signal $a_{j,t}$ provides a presynaptic component but also evokes a basic depolarization of the postsynaptic neuron. The reinforcement signal will provide an additional depolarizing component and the magnitude of this component will determine whether the synapse will be strengthened (long-term potentiation, LTP) or weakened (long-term depression LTD; Pennartz, 1997). If $a_{r,t} > w_{ij,t}$ then LTP will be induced, whereas LTD occurs when $a_{r,t} < w_{ij,t}$. This form of value learning is achieved by a reverse replay of CS1–CS2 sequences in the ensembles A and B, as well as A′ and B′, coupled to a steeply peaking and slowly decaying reinforcement signal broadcast to all cells in A′ and B′.

In a more elaborate version of this model, errors in reward prediction, including surrogate prediction errors (Eq. (10)), can be computed to enable backwards referral of value information as is the case in TD-learning. This variant contains the same sensory-evaluative ensembles and decision network as in Fig. 7, but now the values $V_A(t)$ and $V_B(t)$ are continuously fed from ensembles A′ and B′ not only into the decisional network, but also to the RPM, which differentiates these inputs (i.e., calculates $V_A(t) - V_A(t-1)$, etc.), subsequently adds this difference to the actual reward r_t and emits the result of this computation, the error in reward prediction, towards ensembles A′ and B′ (cf. Fig. 5). Again, the CS1–CS2 sequences are replayed in reverse in the ensembles A and B and A′ and B′, while the error signals are expressed as steeply peaking and decaying transients in, e.g., dopamine release. Note that any form of replay is only possible when an initial, short-term storage of stimulus and reward information has taken place at the time of the original behavioural experience.

The net result of the learning operations sketched above is that $V_A(t)$ and $V_B(t)$ come to represent the discounted reward values of CS1 and CS2, respectively. When these signals are propagated into the decisional network, the selection principles at work within this module, such as a winner-take-all mechanism, will lead the animal to favour a particular action (e.g., ‘approach CS1’) over another one (e.g., ‘approach CS2’), based on the different strengths of $V_A(t)$ and $V_B(t)$.

Admittedly, this model is quite elementary and will need to be elaborated further to account for the richness of findings on intertemporal decision-making. Additional capacities must be attributed to the decisional network to render it functional, such as an ability to suppress or withhold motor actions upon instruction by additional cues. Furthermore, in Fig. 7 decisions have been placed under the control of Pavlovian processes attributing value to initially neutral stimuli, whereas there is extensive behavioural and neurophysiological evidence for separate representation of action values (Dayan and Balleine, 2002; Mulder et al., 2003; Samejima et al., 2005). A more complete model should also be capable of performing additional computations to account for habit-based influences on decision-making and modulation of reward, action and stimulus valuation by the animal’s motivational state. Advancement of the field can be expected particularly when such modelling and simulation efforts will be paired with behavioural, neurophysiological and pharmacological studies on replay phenomena and intertemporal decision-making.

6. Summary and outlook

In this review, we show that virtually all species examined thus far discount temporally proximal events stronger than temporally distant ones. This results in a typical choice pattern that frequently yields suboptimal outcomes and is in violation with several axioms of DUT, including the assumptions of stationary preferences, consistent preference orders and utility maximisation. Although the two crucial parameters influencing intertemporal decisions, delay and reward amount, are processed in many different parts in the brain, the search for a time-varying, hyperbolically discounted value signal has not yet yielded conclusive results. Whereas some studies indicate that the discounted reward value is represented on a single-cell level as an integrated function of reward amount and delay (“common currency”), other experiments suggest it is represented on a population level, and even other studies imply that intertemporal decisions are the result of multiple competing and interacting processes distributed across many different brain regions. Future research needs to address more systematically if, where and how in the brain delay and reward information are integrated into a discounted value representation.

Many of the existing models of classical and operant conditioning are, in their present form, insufficient to explain how time-discounted reward values are updated in the brain, how they are attributed to reward-predicting cues, and how an animal makes its choice. In addition to the possibility of steeply decaying ‘eligibility traces’, we propose a new, simple and biologically plausible model that assumes that task-relevant reward- and CS-features, including the delay between CS and reward, is reversely replayed during rest. We further maintain that a steeply decaying reinforcement signal (e.g., dopamine or glutamate), whose amplitude is scaled to reward amount, coincides with the reversely replayed CS-information, resulting in the association of higher value to temporally proximal cue-representations than to distal events, and to cues predicting larger vs. smaller rewards. This allows an animal to assign delay- and reward-dependent value to the reward-predicting cues, and make its choice between outcomes with different delays. The model does not exclude other mechanisms, embodied by one or some of the other proposed models, to be at work at the same time, performing either supplementary functions or working in concert with reverse-replay operations.

This model makes several predictions that need to be tested in future studies. In addition to further research needed to elucidate general features of reinforcement learning, e.g., the nature and origin of reinforcement signals, it will be necessary to show that

1. CS-reward information is indeed replayed in reversed order during rest while relative delay information is preserved, possibly in a time-compressed manner.
2. Reverse replay happens in more dopaminergic/glutamatergic target regions than only hippocampus, e.g., ventral striatum, amygdala or OFC.
3. The timing of reinforcement signal onset and the onset of replay is coordinated: the replay happens in the awake condition shortly after reward delivery, allowing the reinforcement signal to coincide with the replay of CS-representations.
4. Blocking the transmitter/receptor systems mediating CS-information or the reinforcement signal during the replay-period should abolish learning and learning-dependent preference reversals.

The neuroscience of intertemporal decisions is only at its starting point. But already now, it adds a valuable contribution to the empirical investigation of this type of choice behaviour. We hope that the insights gained in this new field will inform psychology, economics and biology as much as it has benefited from these disciplines.

Acknowledgements

We would like to thank Sabine Windmann, Jeroen Bos, Henrique Cabral, Martin Kocher and four anonymous reviewers for very valuable comments on earlier versions of this manuscript. This work was supported by grants from SenterNovem (BSIK-03053) and the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO-918.46.609).

References

Acheson, A., Farrar, A.M., Patak, M., Hausknecht, K.A., Kieres, A.K., Choi, S., de Wit, H., Richards, J.B., 2006. Nucleus accumbens lesions decrease sensitivity to rapid changes in the delay to reinforcement. Behav. Brain Res. 173, 217–228.

Ainslie, G., 1974. Impulse control in pigeons. J. Exp. Anal. Behav. 21, 485–489.

Ainslie, G., 1975. Specious reward, a behavioral theory of impulsiveness and impulse control. Psychol. Bull. 82, 463–496.

Apicella, P., Ljungberg, T., Scarnati, E., Schultz, W., 1991. Responses to reward in monkey dorsal and ventral striatum. Exp. Brain Res. 85, 491–500.

Arkes, H.R., Ayton, P., 1999. The sunk cost and Concorde effects, Are humans less rational than lower animals? Psychol. Bull. 125, 591–600.

Barraclough, D.J., Conroy, M.L., Lee, D., 2004. Prefrontal cortex and decision making in a mixed-strategy game. Nat. Neurosci. 7, 404–410.

Bassareo, V., DiChiara, G., 1999. Differential responsiveness of dopamine transmission to food-stimuli in nucleus accumbens shell/core compartments. Neuroscience 89, 637–641.

Baxter, M.G., Murray, E.A., 2002. The amygdala and reward. Nat. Rev. Neurosci. 3, 563–573.

Baxter, M.G., Parker, A., Lindner, C.C.C., Izquierdo, A.C., Murray, E.A., 2000. Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. J. Neurosci. 20, 4311–4319.

Bechara, A., Damasio, H., Damasio, A.R., 2000a. Emotion, decision making and the orbitofrontal cortex. Cereb. Cortex 10, 295–307.

Bechara, A., Damasio, H., Tranel, D., Anderson, S.W., 1998. Dissociation of working memory from decision making within the human prefrontal cortex. J. Neurosci. 18, 428–437.

Bechara, A., Tranel, D., Damasio, H., 2000b. Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. Brain 123, 2189–2202.

Bechara, A., Tranel, D., Damasio, H., Damasio, A.R., 1996. Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. Cereb. Cortex 6, 215–225.

Beiser, D.G., Hua, S.E., Houk, J.C., 1997. Network models of the basal ganglia. Curr. Opin. Neurobiol. 7, 185–190.

Bennett, S.M., 2002. Preference reversal and the estimation of indifference points using a fast-adjusting delay procedure with rats. Dissertation. University of Florida.

Benzion, U., Rapoport, A., Yagil, J., 1989. Discount rates inferred from decisions. An experimental study. Manage. Sci. 35, 270–284.

Bernoulli, D., 1954. Exposition of a new theory on the measurement of risk. Econometrica 22, 23–36 (originally published in 1738, translated into English in 1954).

Berns, G.S., Chappelow, J., Cekic, M., Zink, C.F., Pagnoni, G., Martin-Skurski, M.E., 2006. Neurobiological substrates of dread. Science 312, 754–758.

Berridge, K.C., Robinson, T.E., 1998. What is the role of dopamine in reward, hedonic impact, reward learning, or incentive salience? Brain Res. Brain Res. Rev. 28, 309–369.

Berridge, K.C., Robinson, T.E., 2003. Parsing reward. Trends Neurosci. 26, 507–513.

Bi, G.Q., Poo, M.M., 1998. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J. Neurosci. 18, 10464–10472.

Brody, C.D., Hernandez, A., Zainos, A., Romo, R., 2003. Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. Cereb. Cortex 13, 1196–1207.

Buhusi, C.V., Meck, W.H., 2005. What makes us tick? Functional and neural mechanisms of interval timing. Nat. Rev. Neurosci. 6, 755–765.

Buonomano, D.V., Karmarkar, U.R., 2002. How do we tell time? Neuroscientist 8, 42–51.

Cardinal, R.N., 2006. Neural systems implicated in delayed and probabilistic reinforcement. Neural Networks 19, 1277–1301.

Cardinal, R.N., Pennicott, D.R., Sugathapala, C.L., Robbins, T.W., Everitt, B.J., 2001. Impulsive choice induced in rats by lesions of the nucleus accumbens core. Science 292, 2499–2501.

Cardinal, R.N., Robbins, T.W., Everitt, B.J., 2000. The effects of *d*-amphetamine, chlordiazepoxide, alpha-flupenthixol and behavioural manipulations on choice of signalled and unsignalled delayed reinforcement in rats. Psychopharmacology (Berlin) 152, 362–375.

Cardinal, R.N., Howes, N.J., 2005. Effects of lesions of the nucleus accumbens core on choice between small certain rewards and large uncertain rewards in rats. BMC Neurosci. 6, 37.

Chung, S.H., Herrnstein, R.J., 1967. Choice and delay of reinforcement. J. Exp. Anal. Behav. 10, 67–74.

Church, R.M., Meck, W.H., 2003. A concise introduction to scalar timing theory. In: Meck, W.H. (Ed.), Functional and Neural Mechanisms of Interval Timing, vol. 1. CRC Press, Boca Raton, FL, pp. 3–22.

Clayton, N.S., Krebs, J.R., 1995. Memory in food-storing birds: from behaviour to brain. Curr. Opin. Neurobiol. 5, 149–154.

Cragg, S.J., Rice, M.E., 2004. Dancing past the DAT at a DA synapse. Trends Neurosci. 27, 270–277.

Cromwell, H.C., Schultz, W., 2003. Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. J. Neurophysiol. 89, 2823–2838.

Daw, N.D., Doya, K., 2006. The computational neurobiology of learning and reward. Curr. Opin. Neurobiol. 16, 199–204.

Dayan, P., Balleine, B.W., 2002. Reward, motivation, and reinforcement learning. Neuron 36, 285–298.

Diekamp, B., Kalt, T., Güntürkün, O., 2002. Working memory neurons in pigeons. J. Neurosci. 22 (RC210), 1–5.

Dorris, M.C., Glimcher, P.W., 2004. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. Neuron 44, 365–378.

Durstewitz, D., 2003. Self-organizing neural integrator predicts interval times through climbing activity. J. Neurosci. 23, 5342–5353.

Durstewitz, D., 2004. Neural representation of interval time. Neuroreport 15, 745–749.

Estle, S.J., Green, L., Myerson, J., Holt, D.D., 2007. Discounting of monetary and directly consumable rewards. Psychol. Sci. 18, 58–63.

Evenden, J.L., 1999. Varieties of impulsivity. Psychopharmacology (Berlin) 146, 348–361.

Evenden, J.L., Ryan, C.N., 1996. The pharmacology of impulsive behaviour in rats: the effects of drugs on response choice with varying delays of reinforcement. Psychopharmacology (Berlin) 128, 161–170.

Fehr, E., 2002. The economics of impatience. Nature 415, 269–272.

Fiorillo, C.D., Tobler, P.N., Schultz, W., 2003. Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299, 1898–1902.

Fishburn, P.C., Rubinstein, A., 1982. Time preference. Int. Econ. Rev. 23, 677–694.

Foster, D.J., Wilson, M.A., 2006. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature 440, 680–683.

Frederick, S., Loewenstein, G.F., O'Donoghue, T., 2002. Time discounting and time preference: a critical review. J. Econ. Lit. 40, 351–401.

Friedman, M., Savage, L.J., 1948. The utility analysis of choices involving risk. J. Polit. Econ. 56, 279–304.

Friedman, M., Savage, L.J., 1952. The expected-utility hypothesis and the measurability of utility. J. Polit. Econ. 60, 463–474.

Fudenberg, D., Levine, D.K., 2006. A dual-self model of impulse control. Am. Econ. Rev. 96, 1449–1476.

Fuster, J.M., 2000. Executive frontal functions. Exp. Brain Res. 133, 66–70.

Gervasi, N., Hepp, R., Tricoire, L., Zhang, J., Lambolez, B., Paupardin-Tritsch, D., Vincent, P., 2007. Dynamics of protein kinase A signaling at the membrane, in the cytosol, and in the nucleus of neurons in mouse brain slices. J. Neurosci. 27, 2744–2750.

Gibbon, J., 1977. Scalar expectancy theory and Weber's law in animal timing. Psychol. Rev. 84, 279–325.

Glimcher, P.W., 2004. Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics. MIT Press, Cambridge, MA.

Glimcher, P.W., Kable, J., Louie, K., 2007. Neuroeconomic studies of impulsivity: now or just as soon as possible? Am. Econ. Rev. 97, 142–147.

Glimcher, P.W., Rustichini, A., 2004. Neuroeconomics: the consilience of brain and decision. Science 306, 447–452.

Gottfried, J.A., O'Doherty, J., Dolan, R.J., 2002a. Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. J. Neurosci. 22, 10829–10837.

Gottfried, J.A., Deichmann, R., Winston, J.S., Dolan, R.J., 2002b. Functional heterogeneity in human olfactory cortex: an event-related functional magnetic resonance imaging study. J. Neurosci. 22, 10819–10828.

Green, L., Fisher Jr., E.B., Perlow, S., Sherman, L., 1981. Preference reversal and self-control: choice as a function of reward amount and delay. Behav. Anal. Lett. 1, 43–51.

Green, L., Fristoe, N., Myerson, J., 1994. Temporal discounting and preference reversals in choice between delayed outcomes. Psychon. Bull. Rev. 1, 383–389.

Green, L., Myerson, J., 1996. Exponential versus hyperbolic discounting of delayed outcomes: risk and waiting time. Am. Zool. 36, 496–505.

Green, L., Myerson, J., 2004. A discounting framework for choice with delayed and probabilistic rewards. Psychol. Bull. 130, 769–792.

Green, L., Myerson, J., McFadden, E., 1997. Rate of temporal discounting decreases with amount of reward. Mem. Cogn. 25, 715–723.

Gregorios-Pippas, L.V., Tobler, P.N., Schultz, W., 2005. Processing of reward delay and magnitude in the human brain. In: Annual Meeting of the Society for Neuroscience, Program No. 74.6.

Grimm, J.W., See, R.E., 2000. Dissociation of primary and secondary reward-relevant limbic nuclei in an animal model of relapse. Neuropsychopharmacology 22, 473–479.

Grossbard, C.L., Mazur, J.E., 1986. A comparison of delays and ratio requirements in self-control choice. J. Exp. Anal. Behav. 45, 305–315.

Gul, F., Pesendorfer, W., 2001. Temptation and self-control. Econometrica 69, 1403–1435.

Güntürkün, O., 2005. The avian 'prefrontal cortex' and cognition. Curr. Opin. Neurobiol. 15, 686–693.

Hariri, A.R., Brown, S.M., Williamson, D.E., Flory, J.D., de Wit, H., Manuck, S.B., 2006. Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity. J. Neurosci. 26, 13213–13217.

Hartje, W., Poeck, K., 1997. Klinische Neuropsychologie, 3rd edition. Georg Thieme Verlag, Stuttgart.

Hasegawa, R.P., Blitz, A.M., Geller, N.L., Goldberg, M.E., 2000. Neurons in monkey prefrontal cortex that track past or predict future performance. Science 290, 1786–1789.

Hassani, O.K., Cromwell, H.C., Schultz, W., 2001. Influence of expectation of different rewards on behavior-related neuronal activity in the striatum. J. Neurophysiol. 85, 2477–2489.

Hayden, B.Y., Platt, M.L., 2007. Temporal discounting predicts risk sensitivity in rhesus macaques. Curr. Biol. 17, 49–53.

Heath, C., 1995. Escalation and de-escalation of commitment in response to sunk costs: the role of budgeting in mental accounting. Org. Behav. Hum. Decis. Processes 62, 38–54.

Heien, M.L., Johnson, M.A., Wightman, R.M., 2004. Resolving neurotransmitters detected by fast-scan cyclic voltammetry. Anal. Chem. 76, 5697–5704.

Hempel, C.M., Vincent, P., Adams, S.R., Tsien, R.Y., Selverston, A.I., 1996. Spatio-temporal dynamics of cyclic AMP signals in an intact neural circuits. Nature 384, 166–169.

Hestrin, S., 1992. Activation and desensitization of glutamate-activated channels mediating fast excitatory synaptic currents in the visual cortex. Neuron 9, 991–999.

Hikosaka, K., Watanabe, M., 2000. Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. Cereb. Cortex 10, 263–271.

Hikosaka, K., Watanabe, M., 2004. Long- and short-range reward expectancy in the primate orbitofrontal cortex. Eur. J. Neurosci. 19, 1046–1054.

Hinton, S.C., Meck, W.H., 1997. The 'internal clocks' of circadian and interval timing. Endeavour 21, 82–87.

Hodges, C.M., Wolf, L.L., 1981. Optimal foraging in bumblebees: why is nectar left behind in flowers? Behav. Ecol. Sociobiol. 9, 41–44.

Hollerman, J.R., Schultz, W., 1998. Dopamine neurons report an error in the temporal prediction of reward during learning. Nat. Neurosci. 1, 304–309.

Hollerman, J.R., Tremblay, L., Schultz, W., 1998. Influence of reward expectation on behavior-related neuronal activity in primate striatum. J. Neurophysiol. 80, 947–963.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., Camerer, C.F., 2005. Neural systems responding to degrees of uncertainty in human decision-making. Science 310, 1680–1683.

Hwang, J., Kim, S., Lee, D., 2006. Neuronal signals related to delayed reward and its discounted value in the macaque dorsolateral prefrontal cortex. In: Annual Meeting of the Society for Neuroscience, Program No. 71.7.

Isles, A.R., Humby, T., Walters, E., Wilkinson, L.S., 2004. Common genetic effects on variation in impulsivity and activity in mice. J. Neurosci. 24, 6733–6740.

Isles, A.R., Humby, T., Wilkinson, L.S., 2003. Measuring impulsivity in mice using a novel operant delayed reinforcement task: effects of behavioural manipulations and D-amphetamine. Psychopharmacology (Berlin) 170, 376–382.

Ito, M., Asaki, K., 1982. Choice behavior of rats in a concurrent-chains schedule: amount and delay of reinforcement. J. Exp. Anal. Behav. 37, 383–392.

Ivry, R.B., 1996. The representation of temporal information in perception and motor control. Curr. Opin. Neurobiol. 6, 851–857.

Ivry, R.B., 1997. Cerebellar timing systems. Int. Rev. Neurobiol. 41, 555–573.

Ivry, R.B., Spencer, R.M., Zelaznik, H.N., Diedrichsen, J., 2002. The cerebellum and event timing. Ann. N. Y. Acad. Sci. 978, 302–317.

Izawa, E., Aoki, N., Matsushima, T., 2005. Neural correlates of the proximity and quantity of anticipated food rewards in the ventral striatum of domestic chicks. Eur. J. Neurosci. 22, 1502–1512.

Izawa, E., Zachar, G., Yanagihara, S., Matsushima, T., 2003. Localized lesion of caudal part of lobus parolfactorius caused impulsive choice in the domestic chick, evolutionarily conserved function of ventral striatum. J. Neurosci. 23, 1894–1902.

Janssen, P., Shadlen, M.N., 2005. A representation of the hazard rate of elapsed time in macaque area LIP. Nat. Neurosci. 8, 234–241.

Jarvis, E.D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., Medina, L., Paxinos, G., Perkel, D.J., Shimizu, T., Striedter, G., Wild, J.M., Ball, G.F., Dugas-Ford, J., Durand, S.E., Hough, G.E., Husband, S., Kubikova, L., Lee, D.W., Mello, C.V., Powers, A., Siang, C., Smulders, T.V., Wada, K., White, S.A., Yamamoto, K., Yu, J., Reiner, A., Butler, A.B., 2005. Avian brains and a new understanding of vertebrate brain evolution. Nat. Rev. Neurosci. 6, 151–159.

Ji, D., Wilson, M.A., 2007. Coordinated memory replay in the visual cortex and hippocampus during sleep. Nat. Neurosci. 10, 100–107.

Jones, B., Rachlin, H., 2006. Social discounting. Psychol. Sci. 17, 283–286.

Kacelnik, A., Bateson, M., 1997. Risk-sensitivity: crossroads for theories of decision-making. Trends Cogn. Sci. 1, 304–309.

Kacelnik, A., 1997. Normative and descriptive models of decision making: time discounting and risk sensitivity. In: Bock, G.R., Cardew, G. (Eds.), Characterizing Human Psychological Adaptations. Ciba Foundation Symposium, vol. 208. Wiley, Chichester, pp. 51–70.

Kacelnik, A., 2006. Meanings of rationality. In: Hurley, S., Nudds, M. (Eds.), Rational Animals? Oxford University Press, Oxford, pp. 5–43.

Kacelnik, A., Bateson, M., 1996. Risky theories—the effects of variance on foraging decisions. Am. Zool. 36, 402–434.

Kagel, J.H., Green, L., Caraco, T., 1986. When foragers discount the future: constraint or adaptation? Anim. Behav. 34, 271–283.

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47, 263–291.

Kalenscher, T., Diekamp, B., Güntürkün, O., 2003. Neural architecture of choice behaviour in a concurrent interval schedule. Eur. J. Neurosci. 18, 2627–2637.

Kalenscher, T., Güntürkün, O., Calabrese, P., Gehlen, W., Kalt, T., Diekamp, B., 2005a. Neural correlates of a default response in a delayed go/no-go task. J. Exp. Anal. Behav. 84, 521–535.

Kalenscher, T., Windmann, S., Diekamp, B., Rose, J., Güntürkün, O., Colombo, M., 2005b. Single units in the pigeon brain integrate reward amount and time-to-reward in an impulsive choice task. Curr. Biol. 15, 594–602.

Kalenscher, T., Ohmann, T., Güntürkün, O., 2006a. The neuroscience of impulsive and self-controlled decisions. Int. J. Psychophysiol. 62, 203–211.

Kalenscher, T., Ohmann, T., Windmann, S., Freund, N., Güntürkün, O., 2006b. Single forebrain neurons represent interval timing and reward amount during response scheduling. Eur. J. Neurosci. 24, 2923–2931.

Karlsson, N., Gärling, T., Bonini, N., 2005. Escalation of commitment with transparent future outcomes. Exp. Psychol. 52, 67–73.

Keller, B.U., Konnerth, A., Yaari, Y., 1991. Patch clamp analysis of excitatory synaptic currents in granule cells of rat hippocampus. J. Physiol. 435, 275–293.

Keren, G., Roelofsma, P., 1995. Immediacy and certainty in intertemporal choice. Org. Behav. Hum. Decis. Processes 63, 287–297.

Kheramin, S., Body, S., Mobini, S., Ho, M.Y., Velazquez-Martinez, D.N., Bradshaw, C.M., Szabadi, E., Deakin, J.F., Anderson, I.M., 2002. Effects of quinolinic acid-induced lesions of the orbital prefrontal cortex on intertemporal choice: a quantitative analysis. Psychopharmacology (Berlin) 165, 9–17.

Kim, J.N., Shadlen, M.N., 1999. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat. Neurosci. 2, 176–185.

Kirby, K.N., Herrnstein, R.J., 1995. Preference reversals due to myopic discounting of delayed rewards. Psychol. Sci. 6, 83–89.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., Glover, G., 2005. Distributed neural representation of expected value. J. Neurosci. 25, 4806–4812.

Kogut, C.A., 1990. Consumer search behavior and sunk costs. J. Econ. Behav. Org. 14, 381–392.

Kojima, S., Goldman-Rakic, P.S., 1982. Delay-related activity of prefrontal neurons in rhesus monkeys performing delayed response. Brain Res. 248, 43–49.

Kolb, B., Whishaw, I.Q., 2003. Fundamentals of Human Neuropsychology, 5th edition. Worth Publishers, New York.

Komura, Y., Tamura, R., Uwano, T., Nishijo, H., Kaga, K., Ono, T., 2001. Retrospective and prospective coding for predicted reward in the sensory thalamus. Nature 412, 546–549.

Koopmans, T.C., 1960. Stationary ordinal utility and impatience. Econometrica 28, 287–309.

Kröner, S., Güntürkün, O., 1999. Afferent and efferent connections of the caudolateral neostriatum in the pigeon (*Columba livia*): a retro- and anterograde pathway tracing study. J. Comp. Neurol. 407, 228–260.

Kudrimoti, H.S., Barnes, C.A., McNaughton, B.L., 1999. Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and EEG dynamics. J. Neurosci. 19, 4090–4101.

Laibson, D., 1997. Golden eggs and hyperbolic discounting. Quart. J. Econ. 112, 443–477.

Lancaster, K., 1963. An axiomatic theory of consumer time preference. Int. Econ. Rev 4, 221–231.

Lansink, C.S., Goltstein, P., Joosten, R.J.N.M.A., Lankelma, J., McNaughton, B.L., Pennartz, C.M.A., 2007. Coherent reactivation in hippocampal–ventral striatal ensembles during sleep. Soc. Neurosci. Abstr. (Abstract Program number 207.17).

Lavin, A., Nogueira, L., Lapish, C.C., Wightman, R.M., Phillips, P.E., Seamans, J.K., 2005. Mesocortical dopamine neurons operate in distinct temporal domains using multimodal signaling. J. Neurosci. 25, 5013–5023.

Lee, A.K., Wilson, M.A., 2002. Memory of sequential experience in the hippocampus during slow wave sleep. Neuron 36, 1183–1194.

Lee, D., Rushworth, M.F.S., Walton, M.E., Watanabe, M., Sakagami, M., 2007. Functional specialization of the primate frontal cortex during decision making. J. Neurosci. 27, 8170–8173.

Leon, M.I., Shadlen, M.N., 1999. Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. Neuron 24, 415–425.

Leon, M.I., Shadlen, M.N., 2003. Representation of time by neurons in the posterior parietal cortex of the macaque. Neuron 38, 317–327.

Levy, W.B., Steward, O., 1983. Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. Neuroscience 8, 791–797.

Lewis, P.A., Miall, R.C., 2003. Brain activation patterns during measurement of sub- and supra-second intervals. Neuropsychologia 41, 1583–1592.

Lisman, J., Schulman, H., Cline, H., 2002. The molecular basis of CaMKII function in synaptic and behavioural memory. Nat. Rev. Neurosci. 3, 175–190.

Lissek, S., Güntürkün, O., 2003. Dissociation of extinction and behavioral disinhibition: the role of NMDA receptors in the pigeon associative forebrain during extinction. J. Neurosci. 23, 8119–8124.

Loewenstein, G.F., 1987. Anticipation and the valuation of delayed consumption. Econ. J. 97, 666–684.

Loewenstein, G.F., 1988. Frames of mind in intertemporal choice. Manage. Sci. 34, 200–214.

Loewenstein, G.F., 1992. The fall and rise of psychological explanations in the economics of intertemporal choice. In: Loewenstein, G., Elster, J. (Eds.), Choice Over Time. Russell Sage Foundation, New York, pp. 3–34.

Loewenstein, G.F., O'Donoghue, T., 2004. Animal Spirits: Affective and Deliberative Processes in Economic Behavior. Working Paper. Cornell University, Center for Analytic Economics. Available at SSRN: http://ssrn.com/abstract=539843.

Logue, A.W., 1988. Research on self-control: an integrating framework. Behav. Brain Sci. 11, 665–709.

Louie, K., Glimcher, P.W., 2006. Temporal discounting activity in monkey parietal neurons during intertemporal choice. In: Annual Meeting of the Society for Neuroscience, Program No. 605.5.

Majewska, A., Brown, E., Ross, J., Yuste, R., 2000. Mechanisms of calcium decay kinetics in hippocampal spines: role of spine calcium pumps and calcium diffusion through the spine neck in biochemical compartmentalization. J. Neurosci. 20, 1722–1734.

Malaka, R., Hammer, M., 1996. Real-time models of classical conditioning. In: Proceedings of the International Conference on Neural Networks ICNN'96, Washington. IEEE Press, Piscataway, NJ, pp. 768–773.

Markram, H., Lubke, J., Frotscher, M., Sakmann, B., 1997. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science 275, 213–215.

Matell, M.S., Meck, W.H., 2000. Neuropsychological mechanisms of interval timing behavior. Bioessays 22, 94–103.

Mazur, J.E., 1984. Tests of an equivalence rule for fixed and variable reinforcer delays. J. Exp. Psychol. Anim. Behav. Process. 10, 426–436.

Mazur, J.E., 1987. An adjusting procedure for studying delayed reinforcement. In: Commons, M.L., Mazur, J.E., Nevin, J.A., Rachlin, H. (Eds.), Quantitative Analyses of Behavior, vol. 5. The Effect of Delay and of Intervening Events on Reinforcement Value. Erlbaum, Hillsdale, NJ, pp. 55–73.

Mazur, J.E., 1988. Estimation of indifference points with an adjusting-delay procedure. J. Exp. Anal. Behav. 49, 37–47.

Mazur, J.E., 1989. Theories of probabilistic reinforcement. J. Exp. Anal. Behav. 51, 87–99.

McClure, S.M., Ericson, K.M., Laibson, D.I., Loewenstein, G., Cohen, J.D., 2007. Time discounting for primary rewards. J. Neurosci. 27, 5796–5804.

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.

McCoy, A.N., Platt, M.L., 2005. Risk-sensitive neurons in macaque posterior cingulated cortex. Nat. Neurosci. 8, 1220–1227.

McDiarmid, C.G., Rilling, M.E., 1965. Reinforcement delay and reinforcement rate as determinants of schedule preference. Psychon. Sci. 2, 195–196.

Metcalfe, J., Mischel, W., 1999. A hot/cool-system analysis of delay of gratification: dynamics of willpower. Psychol. Rev. 106, 3–19.

Mischel, W., Grusec, J., 1967. Waiting for rewards and punishments: effects of time and probability of choice. J. Pers. Soc. Psychol. 5, 24–31.

Mobini, S., Body, S., Ho, M.Y., Bradshaw, C.M., Szabadi, E., Deakin, J.F., Anderson, I.M., 2002. Effects of lesions of the orbitofrontal cortex on sensitivity to delayed and probabilistic reinforcement. Psychopharmacology (Berlin) 160, 290–298.

Mobini, S., Chiang, T.J., Ho, M.Y., Bradshaw, C.M., Szabadi, E., 2000. Effects of central 5-hydroxytryptamine depletion on sensitivity to delayed and probabilistic reinforcement. Psychopharmacology (Berlin) 152, 390–397.

Mogensen, J., Divac, I., 1982. The prefrontal 'cortex' in the pigeon. Behavioral evidence. Brain Behav. Evol. 21, 60–66.

Mogensen, J., Divac, I., 1993. Behavioural effects of ablation of the pigeon-equivalent of the mammalian prefrontal cortex. Behav. Brain Res. 55, 101–107.

Montague, P.R., Berns, G.S., 2002. Neural economics and the biological substrates of valuation. Neuron 36, 265–284.

Montague, P.R., Dayan, P., Nowlan, S.J., Pouget, A., Sejnowski, T.J., 1993. Using aperiodic reinforcement for directed self-organization during development. In: Hanson, S.J., Cowan, J.D., Giles, C.L. (Eds.), Advances in Neural Information Processing Systems, vol. 5. Morgan Kaufman, San Mateo, CA, pp. 969–977.

Montague, P.R., Dayan, P., Sejnowski, T.J., 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci. 16, 1936–1947.

Moon, H., 2001. Looking forward and looking back: integrating completion and sunk-cost effects within an escalation-of-commitment progress decision. J. Appl. Psychol. 86, 104–113.

Mulder, A.B., Nordquist, R.E., Örgüt, O., Pennartz, C.M.A., 2003. Learning-related changes in response patterns of prefrontal neurons during instrumental conditioning. Behav. Brain Res. 146, 77–88.

Myerson, J., Green, L., 1995. Discounting of delayed rewards: models of individual choice. J. Exp. Anal. Behav. 64, 263–276.

Nadasdy, Z., Hirase, H., Czurko, A., Csicsvari, J., Buzsaki, G., 1999. Replay and time compression of recurring spike sequences in the hippocampus. J. Neurosci. 19, 9497–9507.

Nakazawa, T., Komai, S., Watabe, A.M., Kiyama, Y., Fukaya, M., Arima-Yoshida, F., Horai, R., Sudo, K., Ebine, K., Delawary, M., Goto, J., Umemori, H., Tezuka, T., Iwakura, Y., Watanabe, M., Yamamoto, T., Manabe, T., 2006. NR2B tyrosine phosphorylation modulates fear learning as well as amygdaloid synaptic plasticity. EMBO J. 25, 2867–2877.

Navarro, A.D., Fantino, E., 2005. The sunk cost effect in pigeons and humans. J. Exp. Anal. Behav. 83, 1–13.

Niki, H., Watanabe, M., 1979. Prefrontal and cingulate unit activity during timing behavior in the monkey. Brain Res. 171, 213–224.

Padoa-Schioppa, C., Assad, J.A., 2006. Neurons in the orbitofrontal cortex encode economic value. Nature 441, 223–226.

Parkinson, J.A., Olmstead, M.C., Burns, L.H., Robbins, T.W., Everitt, B.J., 1999. Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive Pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by D-amphetamine. J. Neurosci. 19, 2401–2411.

Pennartz, C.M.A., 1996. The ascending neuromodulatory systems in learning by reinforcement: comparing computational conjectures with experimental findings. Brain Res. Rev. 21, 219–245.

Pennartz, C.M.A., 1997. Reinforcement learning by Hebbian synapses with adaptive thresholds. Neuroscience 81, 303–319.

Pennartz, C.M.A., Lee, E., Verheul, J., Lipa, P., Barnes, C.A., McNaughton, B.L., 2004. The ventral striatum in off-line processing: ensemble reactivation during sleep and modulation by hippocampal ripples. J. Neurosci. 24, 6446–6456.

Pennartz, C.M.A., McNaughton, B.L., Mulder, A.B., 2000. The Glutamate hypothesis of reinforcement learning. Progr. Brain Res. 126, 231–253.

Pennartz, C.M.A., Uylings, H.B.M., Barnes, C.A., McNaughton, B.L., 2002. Memory reactivation and consolidation during sleep: from cellular mechanisms to human performance. Progr. Brain Res. 138, 143–166.

Pickens, C.L., Saddoris, M.P., Setlow, B., Gallagher, M., Holland, P.C., Schoenbaum, G., 2003. Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. J. Neurosci. 23, 11078–11084.

Platt, M.L., Glimcher, P.W., 1999. Neural correlates of decision variables in parietal cortex. Nature 400, 233–238.

Powell, K., 2003. Economy of the mind. PLoS Biol. 1, 312–315.

Prelec, D., Loewenstein, G.F., 1991. Decision making over time and under uncertainty: a common approach. Manage. Sci. 37, 770–786.

Qin, Y.L., McNaughton, B.L., Skaggs, W.E., Barnes, C.A., 1997. Memory reprocessing in corticocortical and hippocampocortical neuronal ensembles. Phil. Trans. R. Soc. Lond. B 352, 1525–1533.

Quintana, J., Fuster, J.M., 1999. From perception to action: temporal integrative functions of prefrontal and parietal neurons. Cereb. Cortex 9, 213–221.

Rachlin, H., 2000. The Science of Self-control. Harvard University Press, Cambridge, MA.

Rachlin, H., Green, L., 1972. Commitment, choice and self control. J. Exp. Anal. Behav. 17, 15–22.

Rachlin, H., Logue, A.W., Gibbon, J., Frankel, M., 1986. Psychol. Rev. Cognition, and behavior in studies of choice 93, 33–45.

Rachlin, H., Raineri, A., Cross, D., 1991. Subjective probability and delay. J. Exp. Anal. Behav. 55, 233–244.

Rainer, G., Miller, E.K., 2002. Time course of object-related neural activity in the primate prefrontal cortex during a short-term memory task. Eur. J. Neurosci. 15, 1244–1254.

Rainer, G., Rao, C., Miller, E.K., 1999. Prospective coding for objects in primate prefrontal cortex. J. Neurosci. 19, 5493–5505.

Raz, J., 1999. Explaining normativity: on rationality and the justification of reason. Ratio 12, 354–379.

Read, D., 2001. Is time-discounting hyperbolic or subadditive? J. Risk Uncertainty 23, 5–32.

Reiner, A., Perkel, D.J., Bruce, L.L., Butler, A.B., Csillag, A., Kuenzel, W., Medina, L., Paxinos, G., Shimizu, T., Striedter, G., Wild, M., Ball, G.F., Durand, S., Güntürkün, O., Lee, D.W., Mello, C.V., Powers, A., White, S.A., Hough, G., Kubikova, L., Smulders, T.V., Wada, K., Dugas-Ford, J., Husband, S., Yamamoto, K., Yu, J., Siang, C., Jarvis, E.D., 2004. Revised nomenclature for avian telencephalon and some related brainstem nuclei. J. Comp. Neurol. 473, 377–414.

Rescorla, R.A., Wagner, A.R., 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A.H., Prokasy, W.F. (Eds.), Classical Conditioning, vol. II. Appleton-Century-Crofts, New York, pp. 64–99.

Reutimann, J., Yakovlev, V., Fusi, S., Senn, W., 2004. Climbing neuronal activity as an event-based cortical representation of time. J. Neurosci. 24, 3295–3303.

Reynolds, B., de Wit, H., Richards, J., 2002. Delay of gratification and delay discounting in rats. Behav. Processes 59, 157–168.

Roelfsema, P.R., 2002. Do neurons predict the future? Science 295, 227.

Roesch, M.R., Olson, C.R., 2004. Neuronal activity related to reward value and motivation in primate frontal cortex. Science 304, 307–310.

Roesch, M.R., Olson, C.R., 2005a. Neuronal activity dependent on anticipated and elapsed delay in macaque prefrontal cortex, frontal and supplementary eye fields, and premotor cortex. J. Neurophysiol. 94, 1469–1497.

Roesch, M.R., Olson, C.R., 2005b. Neuronal activity in primate orbitofrontal cortex reflects the value of time. J. Neurophysiol. 94, 2457–2471.

Roesch, M.R., Taylor, A.R., Schoenbaum, G., 2006. Encoding of time-discounted rewards in orbitofrontal cortex is independent of value representation. Neuron 51, 509–520.

Roesch, M.R., Takahashi, Y., Gugsa, N., Bissonette, G.B., Schoenbaum, G., 2007. Previous cocaine exposure makes rats hypersensitive to both delay and reward magnitude. J. Neurosci. 27, 245–250.

Rohde, K.I.M., 2005. The Hyperbolic Factor: A Measure of Decreasing Impatience. Research Memoranda 044. Maastricht Research School of Economics of Technology and Organization, Maastricht.

Roitman, M.F., Wheeler, R.A., Carelli, R.M., 2005. Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. Neuron 45, 587–597.

Rosati, A.G., Stevens, J.R., Hare, B., Hauser, M.D., 2007. The evolutionary origins of human patience: temporal preferences in chimpanzees, bonobos, and human adults. Curr. Biol. 17, 1663–1668.

Rosenblum, K., Berman, D.E., Hazvi, S., Lamprecht, R., Dudai, Y., 1997. NMDA receptor and the tyrosine phosphorylation of its 2B subunit in taste learning in the rat insular cortex. J. Neurosci. 17, 5129–5135.

Rudebeck, P.H., Walton, M.E., Smyth, A.N., Bannerman, D.M., Rushworth, M.F., 2006. Separate neural pathways process different decision costs. Nat. Neurosci. 9, 1161–1168.

Saddoris, M.P., Gallagher, M., Schoenbaum, G., 2005. Rapid associative encoding in basolateral amygdala depends on connections with orbitofrontal cortex. Neuron 46, 321–331.

Sakurai, Y., Takahashi, S., Inoue, M., 2004. Stimulus duration in working memory is represented by neuronal activity in the monkey prefrontal cortex. Eur. J. Neurosci. 20, 1069–1080.

Samejima, K., Ueda, Y., Doya, K., Kimura, M., 2005. Representation of action-specific reward values in the striatum. Science 310, 1337–1340.

Samuelson, P.A., 1937. A note on measurement of utility. Rev. Econ. Stud. 4, 155–161.

Sanfey, A.G., Loewenstein, G., McClure, S.M., Cohen, J.D., 2006. Neuroeconomics: cross-currents in research on decision-making. Trends Cogn. Sci. 10, 108–116.

Sato, T.R., Schall, J.D., 2003. Effects of stimulus-response compatibility on neural selection in frontal eye field. Neuron 38, 637–648.

Schaubroeck, J., Davis, E., 1994. Prospect theory predictions when escalation is not the only chance to recover sunk costs. Org. Behav. Hum. Decis. Processes 57, 59–82.

Scheuss, V., Yasuda, R., Sobczyk, A., Svoboda, K., 2006. Nonlinear [Ca^{2+}] signaling in dendrites and spines caused by activity-dependent depression of Ca^{2+} extrusion. J. Neurosci. 26, 8183–8194.

Schoenbaum, G., Chiba, A.A., Gallagher, M., 1998. Orbitofrontal Cortex and basolateral amygdala encode expected outcomes during learning. Nat. Neurosci. 1, 155–159.

Schultz, W., Apicella, O., Scarnati, E., Ljungberg, T., 1992. Neuronal activity in monkey ventral striatum related to the expectation of reward. J. Neurosci. 12, 4595–4610.

Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 275, 1593–1599.

Schultz, W., 1997. Dopamine neurons and their role in reward mechanisms. Curr. Opin. Neurobiol. 7, 191–197.

Schultz, W., 1998. Predictive reward signal of dopamine neurons. J. Neurophysiol. 80, 1–27.

Schultz, W., 2002. Getting formal with dopamine and reward. Neuron 36, 241–263.

Schultz, W., 2004. Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. Curr. Opin. Neurobiol. 14, 139–147.

Schweighofer, N., Shishida, K., Han, C.E., Okamoto, Y., Tanaka, S.C., Yamawaki, S., Doya, K., 2006. Humans can adopt optimal discounting strategy under real-time constraints. PLoS Comput. Biol. 2, 1349–1356.

Seamans, J.K., Yang, C.R., 2004. The principal features and mechanisms of dopamine modulation in the prefrontal cortex. Prog. Neurobiol. 74, 1–58.

Seymour, B., O'Doherty, J.P., Dayan, P., Koltzenburg, M., Jones, A.K., Dolan, R.J., Friston, K.J., Frackowiak, R.S., 2004. Temporal difference models describe higher-order learning in humans. Nature 429, 664–667.

Shima, K., Tanji, J., 1998. Role for cingulate motor area cells in voluntary movement selection based on reward. Science 282, 1335–1338.

Shuler, M.G., Bear, M.F., 2006. Reward timing in the primary visual cortex. Science 311, 1606–1609.

Silver, R.A., Traynelis, S.F., Cull-Candy, S.G., 1992. Rapid-time-course miniature and evoked excitatory currents at cerebellar synapses in situ. Nature 355, 163–166.

Skaggs, W.E., McNaughton, B.L., 1996. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. Science 271, 1870–1873.

Snyderman, M., 1987. Prey selection and self-control. In: Commons, M.L., Mazur, J.E., Nevin, J.A., Rachlin, H. (Eds.), Quantitative Analyses of Behavior, the Effect of Delay and of Intervening Events on Reinforcement Value. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 283–308.

Sozou, P.D., 1998. On hyperbolic discounting and uncertain hazard rates. Proc. R. Soc. Lond. B 265, 2015–2020.

Staddon, J.E.R., 2001. Adaptive Dynamics: The Theoretical Analysis of Behavior. MIT/Bradford, Cambridge, MA.

Staddon, J.E.R., Cerutti, D.T., 2003. Operant behavior. Annu. Rev. Psychol. 54, 115–144.

Stephens, D.W., 2002. Discrimination, discounting and impulsivity: a role for an informational constraint. Phil. Trans. R. Soc. Lond. Biol. Sci. 357, 1527–1537.

Stephens, D.W., Anderson, D., 2001. The adaptive value of preference for immediacy: when shortsighted rules have farsighted consequences. Behav. Ecol. 12, 330–339.

Stephens, D.W., Kerr, B., Fernandez-Juricic, E., 2004. Impulsiveness without discounting: the ecological rationality hypothesis. Proc. R. Soc. Lond. B: Biol. Sci. 271, 2459–2465.

Stephens, D.W., Krebs, J.R., 1986. Foraging Theory. Monographs in Behavior and Ecology. Princeton University Press, Princeton, NJ.

Stevens, J.R., Hallinan, E.V., Hauser, M., 2004. The ecology and evolution of patience in two New World monkeys. Biol. Lett. 1, 223–226.

Strotz, R.H., 1955. Myopia and inconsistency in dynamic utility maximization. Rev. Econ. Stud. 23, 165–180.

Sugrue, L.P., Corrado, G.S., Newsome, W.T., 2004. Matching behavior and the representation of value in the parietal cortex. Science 304, 1782–1787.

Suri, R.E., 2001. Anticipatory responses of dopamine neurons and cortical neurons reproduced by internal model. Exp. Brain Res. 140, 234–240.

Sutton, R.S., 1988. Learning to predict by the methods of temporal differences. Mach. Learn. 3, 9–44.

Sutton, R.S., Barto, A.G., 1981. Toward a modern theory of adaptive networks: expectation and prediction. Psychol. Rev. 88, 135–170.

Sutton, R.S., Barto, A.G., 1987. A temporal-difference model of classical conditioning. In: Gabriel, M., Moore, J. (Eds.), Learning and Computational Neuroscience. Proceedings of the Ninth Annual Conference of the Cognitive Science Society, Seattle, WA. MIT Press, Cambridge, MA, pp. 355–378.

Sutton, R.S., Barto, A.G., 1990. Time derivative models of Pavlovian reinforcement. In: Gabriel, M.R., Moore, J.W. (Eds.), Learning and Computational Neuroscience: Foundations of Adaptive Networks. MIT Press, Cambridge, MA, pp. 497–537.

Sweatt, J.D., 2001. The neuronal MAP kinase cascade: a biochemical signal integration system subserving synaptic plasticity and memory. J. Neurochem. 76, 1–10.

Tanaka, S.C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., Yamawaki, S., 2004. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. Nat. Neurosci. 7, 887–893.

Tesauro, G., 1994. TD-gammon, a self-teaching backgammon program, achieves master-level play. Neural Comput. 6, 215–219.

Thaler, R.H., 1981. Some empirical evidence on dynamic inconsistency. Econ. Lett. 8, 201–207.

Thaler, R.H., Shefrin, H.M., 1981. An economic theory of self-control. J. Polit. Econ. 89, 392–406.

Tobler, P.N., Fiorillo, C.D., Schultz, W., 2005. Adaptive coding of reward value by dopamine neurons. Science 307, 1642–1645.

Tobin, H., Logue, A.W., 1994. Self-control across species (*Columba livia*, *Homo sapiens*, and *Rattus norvegicus*). J. Comp. Psychol. 108, 126–133.

Tremblay, L., Schultz, W., 1999. Relative reward preference in primate orbitofrontal cortex. Nature 398, 704–708.

Trepel, C., Fox, C.R., Poldrack, R.A., 2005. Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. Brain Res. Cogn. Brain Res. 23, 34–50.

Van Duuren, E., Nieto-Escámez, F.A., Joosten, R.N.J.M.A., Visser, R., Mulder, A.B., Pennartz, C.M.A., 2007. Neural coding of actual and expected reward magnitude in the orbitofrontal cortex of the rat. Learn. Mem. 14, 613–621.

Von Neumann, J., Morgenstern, O., 1944. Theory of Games and Economic Behavior. Princeton University Press, Princeton, NJ.

Voorn, P., Vanderschuuren, L.J., Groenewegen, H.J., Robbins, T.W., Pennartz, C.M.A., 2004. Putting a spin on the dorsal–ventral divide of the striatum. Trends Neurosci. 27, 468–474.

Wallis, J.D., Miller, E.K., 2003. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. Eur. J. Neurosci. 18, 2069–2081.

Watanabe, M., 1996. Reward expectancy in primate prefrontal neurons. Nature 382, 629–632.

Winstanley, C.A., Theobald, D.E., Cardinal, R.N., Robbins, T.W., 2004. Contrasting roles of basolateral amygdala and orbitofrontal cortex in impulsive choice. J. Neurosci. 24, 4718–4722.

Winstanley, C.A., Baunez, C., Theobald, D.E., Robbins, T.W., 2005. Lesions to the subthalamic nucleus decrease impulsive choice but impair autoshaping in rats: the importance of the basal ganglia in Pavlovian conditioning and impulse control. Eur. J. Neurosci. 21, 3107–3116.

Winstanley, C.A., Theobald, D.E., Dalley, J.W., Cardinal, R.N., Robbins, T.W., 2006. Double dissociation between serotonergic and dopaminergic modulation of medial prefrontal and orbitofrontal cortex during a test of impulsive choice. Cereb. Cortex 16, 106–114.

Wittmann, M., Leland, D.S., Paulus, M.P., 2007. Time and decision making: differential contribution of the posterior insular cortex and the striatum during a delay discounting task. Exp. Brain Res. 179, 643–653.

Yanagihara, S., Izawa, E., Koga, K., Matsushima, T., 2001. Reward-related neuronal activities in basal ganglia of domestic chicks. Neuroreport 12, 1431–1435.

Yang, T., Shadlen, M.N., 2007. Probabilistic reasoning by neurons. Nature 447, 1075–1080.

Yi, R., de la Piedad, X., Bickel, W.K., 2006. The combined effects of delay and probability discounting. Behav. Processes 73, 149–155.