The Temporal Derivative of Expected Utility: A Neural Mechanism for Dynamic Decision-making

Xian Zhang[*,a,d] and **Joy Hirsch**[a,b,c,d]

[a]Department of Psychology, Columbia University

[b]Department of Neuroscience, Columbia University Medical Center

[c]Department of Radiology, Columbia University Medical Center

[d]Department of fMRI Research Center, Columbia University Medical Center

Abstract

Real world tasks involving moving targets, such as driving a vehicle, are performed based on continuous decisions thought to depend upon the temporal derivative of the expected utility ($V/$ t), where the expected utility (V) is the effective value of a future reward. However, those neural mechanisms that underlie dynamic decision-making are not well understood. This study investigates human neural correlates of both V and $V/$ t using fMRI and a novel experimental paradigm based on a pursuit-evasion game optimized to isolate components of dynamic decision processes. Our behavioral data show that players of the pursuit-evasion game adopt an exponential discounting function, supporting the expected utility theory. The continuous functions of V and $V/$ t were derived from the behavioral data and applied as regressors in fMRI analysis, enabling temporal resolution that exceeded the sampling rate of image acquisition, hyper-temporal resolution, by taking advantage of numerous trials that provide rich and independent manipulation of those variables. V and $V/$ t were each associated with distinct neural activity. Specifically, $V/$ t was associated with anterior and posterior cingulate cortices, superior parietal lobule, and ventral pallidum, whereas V was primarily associated with supplementary motor, pre and post central gyri, cerebellum, and thalamus. The association between the $V/$ t and brain regions previously related to decision-making is consistent with the primary role of the temporal derivative of expected utility in dynamic decision-making.

Keywords

Expected utility; temporal derivative of expected utility; dynamic game; fMRI and decision-making

---

[*]Corresponding author at: fMRI Research center of Columbia University Medical Center, 710 W. 168th Street, New York, NY 10032, xz63@columbia.edu (Zhang, X.); Joyhirsch@yahoo.com (Hirsch, J).

Introduction

Expected utility (V) is the product of the probability and subjective utility of a goal. Initiated by Bernoulli (Bernoulli, 1738) and formalized by Morgenstern and von Neumann (Morgenstern and Von Neumann, 1944), expected utility has been a central concept in decision science. This paper investigates the role of temporal change in V, or its temporal derivative $(\partial V/\partial t)$ during decision-making. Many real-world tasks can be modeled as a dynamic game, such as a pursuit-evasion game, where decisions and actions are made continuously, not only responding to, but also altering, the game state. According to dynamic decision theory (Isaacs, 1965), one of the most important decision principles in a dynamic game is maximization of $\partial V/\partial t$ (Appendix 1). The importance of temporal change in V has also been implicated in prominent theories such as the temporal difference algorithm (TD) (Sutton and Barto, 1987), and the prediction error theory (Schultz et al., 1997). Further, single unit data have shown that the activation pattern of some dopamine neurons are similar to $\partial V/\partial t$ (Fig. 1 of (Schultz et al., 1997). It has been hypothesized (Fig.4 of (Schultz et al., 1997), but not demonstrated, that the temporal derivative of utility is coded in the nervous system as a mechanism to enable immediate and reflexive responses. This paper presents a paradigm that allows both V and $\partial V/\partial t$ to be quantified and their neural correlates observed.

Currently, decision-making is most commonly studied with static paradigms, such as a conventional event-related design, where each trial consists of discrete events such as a stimulus, a response and a reward event. Those paradigms are "static" because they consist of discrete decisions where the continuous $\partial V/\partial t$ is not well defined (Basar and Olsder, 1999). Due to the limited temporal resolution of functional magnetic imaging (fMRI) and the absence of independent manipulations of V and $\partial V/\partial t$, it is difficult to separate $\partial V/\partial t$ from V for human subject studies using conventional paradigms. On the other hand, single unit recordings not only describe a variety of temporal profiles of neuronal action potentials but also allow a measure of the neural activity related to the expected utility, prior to the receiving of the reward, and the response to the actual reward separately. However, single unit recordings are invasive and only sample a few neurons and thus do not provide the description of the global neural networks related to the V, $\partial V/\partial t$ and reward.

To meet those challenges, we adopted a pursuit-evasion game, the classic game that was used for developing the dynamic game theory (Isaacs, 1965), to determine both V and $\partial V/\partial t$ as continuous functions of the game states, as well as the capture event as an impulse function. The underlying neural activities associated with those dynamic variables were isolated using fMRI. Our hypothesis is that brain activity related to dynamic decision-making would be correlated with $\partial V/\partial t$ and distinguishable from neural activity associated with V.

Materials and Methods

Task

The task was a continuous pursuit-evasion video game played during fMRI scans. The game was modeled after the familiar Pacman game and subjects aimed to collect 1-point and 2-

point rewards and to avoid 2-point losses, which were treated as gaining negative 2 points. All the characters moved in continuous game space. The character corresponding to the "pellet" as in the Pacman game could also move away from a predator. This computer game adopted a first person's viewpoint. In other words, the avatar of the player was always shown at the center of the monitor. See Appendix 2.1-2.4 for the rules and detailed description of the game. The avatar was controlled by the player with actions such as turning up, down, left and right using an MRI compatible track ball, while other characters were controlled by the game program. The goal of the player was to accumulate a maximum number of points, the unit of utility, by capturing targets. Both the controls and the movements of all the game characters, including the avatar, were modeled as a vehicle with constant speed and minimum turning radius. The speed of a pursuer was 15% greater than that of an evader. Randomly, the game program generated perturbation to the decisions of game characters, modeled as the effect of random wind gusts as on a sail-boat (Appendix 2.5). Such perturbation is important. First, for the player, it simulates the unpredictable nature of the environment. Second, it simulates the random strategy of a computer character. Finally and most importantly, without the perturbation, every pursuit process would be identical, trivially easy and thus failing to provide a rich and independent manipulation of V and $\partial V/\partial t$. In fact, without perturbation, V and $\partial V/\partial t$ would be highly correlated because V is an exponential function.

The continuous game can be divided into two basic phases. First, the avatar faces two targets with comparable expected utility and the player performs a target-choice task (Fig. 1A). Second, when there is no comparable alternative target, the player pursues the chosen target. Fig. 1B represents a dynamic pursuit process in the 2-dimensional game space, where the time is coded in color. V, for the first approximation, depends on the distance between the avatar and the target (Fig. 1C), as well as the value of the target. In summery, this game represents a departure from a conventional trial structure where each event is isolated. Rather, the continuous acquisition of V enables a disassociation among V, $\partial V/\partial t$ and the capture event.

Subjects

Subjects were recruited from the Columbia University Medical Center community and provided written informed consent according to the guidelines of the local institutional review board. Twenty-one subjects (7 females, age=26±8 years) volunteered for the study and participated without compensation. They were given an instruction (Appendix 2.6) and practiced in the scanner prior to image acquisition.

fMRI scan parameters and analysis

To reduce the susceptibility-induced artifacts, we employed a spiral-in/out sequence for the fMRI data acquisition (TE=36ms, TR=2000ms, FA=84, FOV=21cm, Volxel size=3×3×4 mm)(Glover and Thomason, 2004). Functional magnetic resonance imaging (fMRI) images were analyzed using the General linear model for individual data and the mixed effect analysis for the group data using the FSL software package (Smith et al., 2004).

fMRI regressors

V of the entire game state as defined by Eq. 1 was applied to the target that the player is currently pursuing. Potential targets other than the immediate one were ignored because the game was designed so that over 97% of time, the expected utility values of other potential targets were less than half of the value of the current target. $\partial V/\partial t$ was the temporal derivative of V. In addition, following the standard event related paradigm, a capture event was defined as an impulse function at the time of the capture with the magnitude as the value of the target (Fig. 2). V, $\partial V/\partial t$ and capture event were employed as major regressors for a general linear model analysis.

Note that, the captured event in the game differs from the reward event in previous studies, where once chosen, the target will be obtained with certain probability and without further effort. The merit of the conventional approach is that the utility for the reward event is clearly defined, as the product of value and probability of the target. In our game, there is a substantial chance that the target may escape the pursuit in spite of the effort. In other words, the brain activity shown associated with the capture event could also include the activity that previously associated with the prediction error, because a capture event provides information that was uncertain before.

Motion correction and global mean were included as confounding regressors along with the major regressors, including V, $\partial V/\partial t$ and the capturing event described in following sections. In addition, 30% of the capturing events were associated with unexpected outcomes. For example, capturing a 1-point token could result in either a gain of 2 points or a gain of 0 points (see Appendix 3 and Table A3 for a detailed description and Fig. A2 for the fMRI result). Both the capture event and the unexpected payoff regressors included all types of targets, combining both positive (capturing a target) and negative (being captured by the opponent) events.

All regressors were convoluted with the canonical hemodynamic response function, except for the motion correction parameters and the global mean. The orthogonality among all the regressors was confirmed using a correlation measure (Hare et al., 2008). The largest correlation coefficient among all the regressors in this task was 0.03, indicating the independence of the variables for the fMRI analysis.

Task description and theoretical basis for determining the function of expected utility

The role of $\partial V/\partial t$ is illustrated with a typical epoch in a pursuit-evasion game in Fig. 2. During the pursuit process, a sudden change in the direction of movement of the target requires the pursuer, the avatar, to make a decision. Note that, at time $T3$, the direction of movement was non-optimal because the avatar failed to respond to the change. In contrast, at time $T2$, the avatar's direction of movement was optimal. However, the expected utility at $T3$, $V(T3)$, is positive and even greater than $V(T2)$. In contrast, $\partial V(T2)/\partial t$ is positive and $\partial V(T3)/\partial t$ is negative, corresponding to whether the direction of the movement is optimal. Such a situation illustrates the dissociation between V and $\partial V/\partial t$. In addition, Fig. 2 also shows that the function of V, $\partial V/\partial t$ and capturing are distinct to each other, confirming that a pursuit-evasion game may enable us to separate the brain activity specific to those

components of decision-making. In summary, a neuron that codes V cannot make the decision that, at $T3$, it needs to change the direction of movement. Nor can it tell the difference when two targets with the same value and same distance are moving at different speeds. In contrast, coding V/t provides a natural and effective way to utilize game state information.

To understand the function of V for such a complex game, methods from dynamic game theory and microeconomics were combined. First, the game scene was segmented into epochs that ended with a capture. Second, the game state, i.e. the distances between the avatar and the targets, and the action, i.e. the direction control, were converted into a continuous function of V. Since the distance is by and large linearly related to the temporal delay for the reward, we assume that the function V of distance is similar to a temporal discounting function.

Estimating the form and the parameters of the discounting function

Since the game is engineered in the way that random events occurred independently of each other, the rational function of V is expected to be an exponential discounting function (Samuelson, 1937), where the time variable is replaced by distance:

$$V(d, r) = r \times e^{-d/d_0} \quad \text{(Eq. 1)}$$

where d is the distance measured at a time, r is the value of the target and d_0, the discounting constant, is a parameter to be determined. This exponential function (Eq. 1) will be qualitatively compared with a general hyperbolic function (Eq. 2) commonly used in previous decision science research (Loewenstein and Prelec, 1992; Richards et al., 1997; Scholten and Read, 2010):

$$V(d, r) = r \times (1 + \alpha \times d)^{-\beta/\alpha} \quad \text{(Eq. 2)}$$

where α and β are positive parameters to be determined.

The decision function

Although dynamic game theory has established the importance of the temporal derivative of the game state in decision-making, the real time decision function for human subjects in a pursuit-evasion game, to our knowledge, has not been described. Here the real time decision function, which consists of both perception and decision components, is defined as the mathematical operation that transforms input information into the action, in real time.

The game data include the x and y coordinates of all the characters (See Fig. 1B for an example). Both the input information and the avatar's action can be extracted from those data. In this game all characters moved at a constant speed. The only action was the change in the direction of the movement. Therefore, we focus on the direction aspect of the

information. Other information, such as the distance, although related to the expected utility, is irrelevant for a specific action. For example, regardless of the distance, if the target is at one o'clock, the player moves the avatar to that direction. The direction of movement (*Direction*) of the avatar is obtained as:

$$\text{Direction}_t = \text{cart2pol}\,(X_{t+1} - X_t \; Y_{t+1} - Y_t)$$

where cart2pol is the mathematical function transforming a Cartesian coordinate into an angle, t indicates a point in time, *t+1* and *t* are 10 ms apart, and *X, Y* are the coordinates of the avatar in the 2-dimensional game world. The action of the player is measured as the change of *Direction*:

$$\text{Action}_t = \text{Direction}_{t+1} - \text{Direction}_t$$

Note that, *Action$_t$* is the result of a decision responding to information received at some time prior. The direction of location between the avatar and the target (*Angle*) is defined as:

$$\text{Angle}_t = \text{Cart2pol}\,(x_t - X_t, \; y_t - Y_t)$$

where *x* and *y* are the coordinates of the target. The difference between *Direction* and *Angle*, referred to as Δ*Direction*, can be defined as:

$$\Delta\text{Direction}_t = \text{Angle}_t - \text{Direction}_t$$

(Fig. 3A)

For example, if the avatar is moving toward three o'clock, while the target is at one o'clock, Δ*Direction$_t$* is two o'clock or 60°. According to dynamic game theory, the temporal derivative of the game state plays an important role in decision-making. If Δ*Direction* is an aspect of the game state, its temporal derivative (∂Δ*Direction*) can be defined as:

$$\partial\Delta\text{Direction}_t = \Delta\text{Direction}_{t+1} - \Delta\text{Direction}_t$$

(Fig. 3A)

Based on conventional static decision theory, the optimal action should maximize the expected utility (*V*) by minimizing the distance. The most intuitive strategy is to make Δ*Direction* zero, or *Action*= Δ*Direction*, so that the avatar will be moving toward the target. This strategy is optimal when the target is stationary. However, in a dynamic pursuit process, where the target is evading, moving toward the current target position is often not effective. Instead, one should perceive the direction change of the target (∂Δ*Direction*) and make a

move toward the anticipated position. Therefore, dynamic game theory emphasizes the importance of the temporal derivative of the game state. We hypothesize that the decision function should consist of two distinct components, one for ΔDirection and the other for $\partial\Delta$Direction, referred to as *decision*Δ and *decision*$\partial\Delta$. Since there is a time delay between the input information and the action due to reaction time, we use the convolution method to study those decision mechanisms:

$$\text{Action}=\text{decision}\Delta \ast \Delta\text{Direction}+\text{decision}\partial\Delta \ast \partial\Delta\text{Direction}$$

where * is the symbol of convolution. For one trial, *Action,* Δ*Direction* and $\partial\Delta$*Direction* are time series of 150 data points obtained at the final 1.5 second of the pursuit process, when both the player and the target are highly engaged in the process and other characters in the game are irrelevant. A simple way to obtain both decision components, *decision*Δ and *decision*$\partial\Delta$ is to assume that Δ*Direction* and $\partial\Delta$*Direction* are approximately independent of each other:

$$\text{decision}\Delta=\text{Action}\,(\ast-1)\,\Delta\text{Direction}$$
$$\text{decision}\partial\Delta=\text{Action}\,(\ast-1)\,\partial\Delta\text{Direction},$$

where (*-1) is the symbol for de-convolution.

Results

Expected utility function determined by behavioral data

In this game, the player selected a target and controlled the direction of movement for pursuit and capture the target, referred to as a token. *Pursuit-choice* denotes the object toward which the avatar is moving. Examples of *pursuit-choices* between a 1-point (the unit of utility in the game) token and a 2-point token for a typical subject are shown in Fig. 4A, and the group averaged data are shown in Fig. 4B. Note that the decision boundary (the gray area in Fig. 4B) aligns with the prediction of the exponential discounting function (green dashed line), for being parallel to the equal distance (black 45°) line. In other words, when the 1-point token and the 2-point token were chosen with equal probability, the difference between the distances (δ) is constant regardless the average distances for both. Data show that δ is approximately constant regardless the average distances for both targets for all types of decisions in the game (Fig. 4C, see Appendix 2.4 for the details for those choices). Translating distance into temporal delay, this result can be characterized as dynamic consistency, the signature of an exponential discounting function (Schweighofer et al., 2006; Strotz, 1956) (Eq. 1), in agreement with previous findings (see (Schweighofer et al., 2006) or Appendix 4 for a detailed inference). We combined data from all types of decisions in the game to compute the parameter d_0 individually for each subject. Due to the narrow distribution of d_0, the relationship between the performance and the discounting constant is by and large independent for our data (Appendix 5).

The decision function and the temporal derivative of the game state

Fig. 3B shows both *decision*   (red) and *decision*   (blue) based on combined data from all subjects and runs. Both *decision*   and *decision*   contain a significant signal, indicated by the relative magnitude of the peak at 200ms vs. the base line noise. In other words, players apparently encode both   *Direction* and   *Direction* and, after a 200ms delay, probably due to the reaction time, react following those inputs. In particular, our analysis shows that *decision*   is more prominent than *decision*   during the final 1.5 seconds of pursuit processes, consistent with dynamic game theory. Note that the decision functions are narrow peaks, while the reaction time is expected to have a broad distribution. This is the result of the continuous and smooth nature of our input information and limitation of the de-convolution method. The narrow peak indicates the median reaction time.

  Direction and   *Direction* represent a game state and the temporal derivative, and are assumed to be related to the *V* in a complex way. In subsequent analysis, we study the neural correlates for *V* and   *V/ t* calculated directly with distances.

Neural correlates for V and   V/ t

The regressor for *V* co-varied with activity in cerebellum, brainstem, thalamus, supplementary motor area (SMA), pre and postcentral gyri, regions often associated with motor functions (Fig. 5A, table 1A). The regressor for   *V/ t* co-varied with activity in anterior and posterior cingulate cortices (ACC & PCC) and ventral pallidum (VP) (Fig. 5B and table 1B), regions that have previously been related to decision-making and executive functions (McClure et al., 2004; Sarinopoulos et al., 2010). The neural activity correlated with the capture event was found in caudate-putaman and subcallosal cortex, which includes nucleus accumbens (NAC) and the caudate nucleus (CN) (Fig. 5C, table 1C), regions that are part of the ventral striatum and previously related to reward and learning (McNamara et al., 2002). For example, it has been reported that CN showed deactivation in response to failure in a shooting game (Mathiak et al., 2011).

Discussion

The relationships between the choice and the distance of the targets during the games were best described by an exponential function used to model the relationship between the expected utility and the temporal delay and thus, are consistent with the classic expected utility theory and previous observations. This finding does not necessarily contradict more frequently reported results supporting a hyperbolic discounting function. In our game, as well as in real world, random events, such as the emergence of either an unexpected better choice or an unexpected danger; can render a chosen target irrelevant. Therefore, the effective value of a target decreases exponentially with time. In contrast, in an inter-temporal choice paradigm, even the most delayed target has a greater than zero chance to be obtained. As a consequence, along the time dimension, the expected utility asymptotes to a fixed value and thus assumes a hyperbolic function. Therefore, it is possible that the difference in the form of the discounting function may reflect similar decision principals applied to different situations.

The dynamic game paradigm and the behavioral data that are used for generating regressors allow dissociation of important components of decision-making such as V, $\partial V/\partial t$ and capturing events. The major finding in this study is that a neural system consisting of ACC, PCC, SPL, and VP is related with $\partial V/\partial t$, the basis for dynamic decision-making. In other words, those regions are correlated with the temporal derivative rather than the value of expected utility. Such a distinction cannot be made with a conventional experimental paradigm because, for a trial lasting a few seconds, the neural activity related with both V and $\partial V/\partial t$ occurs at essentially the same time for fMRI. On the other hand, single cell recordings, with high temporal resolution, have shown that decision-making related neurons have a phasic response pattern (Moncayo et al., 2000; Opris et al., 2009; Schultz et al., 1997), consistent with a derivative operation. In summary, our results provide additional evidence in support of a neural mechanism related to the derivative of expected utility using fMRI data and human subjects.

Although it is frequently reported that ventral striatum, ACC, PCC and other frontal areas are correlated with V in studies using static decision-making tasks (Kable and Glimcher, 2007; McClure et al., 2004; O'Doherty et al., 2007; Pine et al., 2009), our observation, based on the dissociation of V and $\partial V/\partial t$, add granularity to those findings. The pattern of the activity related to the regressor V was mainly associated with motor related regions. A similar pattern has also been reported with monkey single unit recordings, where the activity of SMA and motor cortex was enhanced when the primate either anticipated a short delay or experienced a long elapsed delay (Roesch and Olson, 2005; Watanabe, 1986) and similar findings were also reported in a human fMRI study (Wunderlich et al., 2009).

The pursuit-evasion video game employed in this study is intended to elicit complex, real-world dynamic decision-making. This departure from the conventional approach to study decision-making as a single input-output event enables the separation of components of a decision-making process. Behavioral and fMRI results suggest that dynamic decision-making is based both upon the exponential discounted utility (V) and its temporal derivative ($\partial V/\partial t$), consistent with dynamic game theory as described in mathematics. The separation between factors of decision-making advances our understanding of dynamic decision-making and understanding the neural activity during complex and dynamic real world situations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Appendix

1. Dynamic game theory

Any dynamic system, including a dynamic game, can be generally described with a differential equation for simplicity and for reducing computational cost:

$$dx/dt = f(t, x(t), u^1(t), u^2(t))$$

where x is the game state or the distance between the pursuer and evader for a pursuit-evasion game. dx/dt is the temporal derivative of x, u^1 and u^2 are the inputs to the system or, for a game, the actions of player 1 and player 2, respectively. Since expected utility (V) is a function related to x, we have:

$$\frac{\partial v}{\partial t} = \frac{\partial v}{\partial x} \times \frac{dx}{dt} = \frac{\partial v}{\partial x} \times f(t, x(t), u^1(t), u^2(t))$$

According to dynamic game theory, maximization *of V/ t* is an important principal for dynamic decision-making (Issacs, 1965).

2. The pursuit-evasion game

2.1 Introduction of the game

In the manuscript, the game state was simplified as the expected utility. However, to engage the subject and to elicit responses resembling those in real world, the task was made to be a complex, realistic video game. This game involves many parameters such as the speed of the evader and a pursuer, the minimum turning radius of any character and the random perturbation. Those parameters were determined with pilot studies to make the game realistic and also possible to analyze.

The game placed the subject in a pursuit-evasion scenario similar to the popular PACMAN game in a two-dimensional world that measures 2000 by 1500 pixels around the avatar of the player. There were 8 characters living in this world: the avatar of the player, the computer opponent, three 1-point tokens and three 2-point tokens (Fig. A1). If any one was either captured or ran outside of the boundary of the world, it would be relocated at a random place in the world. The game was presented with a first person point of view, i.e. the avatar of the player was always shown at the center regardless of its actual location. Therefore, only the relative motion is represented. In dynamic game theory, this presentation of game space is referred to as the "reduced space" (Basar and Olsder, 1999). Only the characters inside the 800×500-pixel part of monitor were visible to the subject. The game state can be described as the coordinates of each of the characters, updated and sampled every 10ms.

Every character moves with a constant speed and the only action is changing in direction. The pursuer is 10% faster than the evader. The game program updates the directions of movement every 200 ms.

2.2 Direction of movement

The subject controlled the direction of the avatar with an MRI compatible trackball to either pursue targets or evade opponent. The screen was divided into 4 regions, left, right, upper and lower, by two invisible 45° lines passing through the center. For example, the direction of movement of the avatar was gradually changed into left when the cursor was inside the "left region". The direction of movement cannot be changed instantaneously; for example, a 90° change in direction takes 3 steps or 600ms, simulating the minimum turning radius of any vehicle.

2.3 Computer opponent's behavior

The opponent calculated the utility of each object based on the formula $V(d, r) = r \times e^{d/10}$ and pursuit the object with maximum V. A token was also controlled by the program. A token can try to escape an approaching pursuer. However it only detects a pursuer within the distance less than 120 pixels.

2.4 Conditions and the payoff tables

There were two basic conditions, called red and blue conditions, alternating every 30 seconds with no gap. In the red condition, the computer opponent was called the red opponent and moved faster than the avatar. In the blue condition, the computer opponent was called the blue opponent and moved slower than the avatar. The outcomes of capture events are listed in Table A1&2. To make the game more attractive and enhance the effect of reward and punishment, the avatar showed an emotional expression for 1 second after any score event (Table A1&A2).

Token-capturing events and the scores for both conditions.

Token capture	No event	1-point	2-point
Score for both players	0	1	2
Avatar Expression			

For both the player and the computer opponent, a token capture was rewarded with 1 point or 2 points depending on the type of the token.

Opponent-colliding events and the scores for both the player and the computer opponent.

Condition and the icon for the opponent:	Red	Blue
Player score	-2	2
Opponent score	2	-2
Avatar Expression		

2.5 Perturbation of movement

To simulate the unpredictable factor in a realistic environment, the directions of the avatar and the opponent could be altered randomly by the game at a probability of 10%. The direction of each token could be altered randomly at a probability of 15%. In other words, the tokens made more mistakes than the avatar and therefore, were easier to be captured. Again, this parameter, along with other parameters, was tuned so that the game is both interesting and analyzable. For a token, when not being pursued, the perturbation results in a random walk. For any character moving toward a desired direction, each perturbation will not have a significant effect because it only lasts for 100ms and will be corrected. However, when perturbations occur in a burst, the subjective feeling of the effect is very similar to a wind gust on a sailboat.

2.6 Instruction to subjects

Today you will be participating in a functional imaging study where we are interested in seeing how your brain responds to certain aspects of a computer game. Your job will be to focus on playing the game. Please try not to move your head.

Rules of the game—Your avatar is the face in the center of the screen. Your goal is to navigate around the environment and collect coins. There are single coins and double coins and as you collect more you will notice the red meter around your avatar building up. Once you have collected a total of 10 points, you will receive a bonus reward that will show up in the bottom left of the screen.

Another aspect to this game is the challenger. There is a single challenger that is also collecting coins. Half of the time, this challenger is red and if he captures you, you may lose up to 2 points. The other half of the time the challenger will be blue. In these cases, you are able to chase him and may earn up to 2 points for capturing him. Watch out because he can change back to red without warning.

The way to control your character is with the trackball. This mouse will show up as a green arrow. The green arrow indicates the change of your direction and not the direction of your motion. The best way to control your movements is to keep the arrow near your character's face. You may notice that sometimes your character will move in ways that is not exactly as

you intended. In fact, he may move in a totally new direction. This is part of the difficulty of the game, so try your best to constantly control the direction of movement.

While earning or losing points through pursuing tokens and the blue competitor, or running away from the red competitor, you may be surprised to see that the actual value is sometimes different from what you expect. For example, your points could be doubled or you may even get zero points. The actual point total for each coin will show up on the screen after you capture it and the facial expression will correspond to how happy or disappointed your character is.

About the fMRI scan—Today you will be completing 5 total runs. In the first 4 runs, you will play the game and each will last 8 minutes. It is very important that you keep your head still throughout the entire 8 minutes. In between runs there will be a short break where you can move slightly, but it is still important to keep your head and body in the same position during this time. After completing these 4 runs, there will be one final scan where you will not be playing the game or viewing anything on the screen. This is to get a high-resolution picture of your brain, so you can close your eyes and rest, but it is still important to not fall asleep and to not move your head or body.

3. Unexpected payoffs

On 70% of the trials, the subject was rewarded as expected. On the remaining 30% of the trials, the rewards for 15% were higher and 15% were lower than the expected (Table A3, A & B).

The unexpected payoff regressor was defined as an impulse function. The magnitude is either 1 for cases that the actual value was greater than the expected value or -1 for the cases that the actual value was smaller than the expected value.

Table A3

A. Unexpected payoff regressor.								
target	2-point token			1-point token			Red opponent	
outcome	2	4	0	1	2	0	-2	0
Chance (%)	70	15	15	70	15	15	70	30
Unexpected	0	1	-1	0	1	-1	0	1
Avata expression For unexpected payoff								

Unexpected payoff summery		
target	Any token	Any opponent

Unexpected payoff summery						
Chance (%)	70	15	15	70	15	15
Unexpected	0	1	-1	0	1	-1

Consistent with previous findings (eg. Hare et al., 2008), the activity of the ventral striatum was correlated with unexpected payoffs (p value < 0.01) (Fig. A2). However, the activity did not exceed the multiple comparison corrected significance level. Note that, the unexpected payoff regressor, although similar, differs from the prediction error in previous studies. The magnitude of the neural activity related to the prediction error is negatively correlated with the probability for the occurrence of the event. In other words, prediction error is related to new information. In our study, the uncertainty for capturing a target is high, regardless of the final payoff. Therefore, the modification of the payoff for a target is much less important than whether the target can be captured at all.

4. Comparison between the exponential vs. hyperbolic discounting function

Let the value of the 2-point token be 2 and the value of the 1-point token be 1. Based on Eq. 1, we have:

$$V_1 = r \times e^{-d/d_0}$$
$$V_2 = 2 \times e^{-d/d_0}, \; V_1 = 1 \times e^{-(d+\Delta)1/d_0}$$

At the decision boundary, i.e. the gray area in Fig. 4B, two types of targets should have identical discounted utility, because subjects chose them with equal probability.

$$V_1 = V_2$$
$$2 \times e^{-d/d_0} = 1 \times e^{-(d+\Delta)1/d_0}$$
$$\ln 2 - d/d0 == (d+\Delta)/d0$$
$$\Delta = d0 * \ln 2$$

In other words, that exponential discounting function requires   be a constant. For example, if the average   is about 14 pixels (Fig. 4), as the data indicate, the average d_0 is 21 pixels for Eq. 1.

However, for a general hyperbolic function (Eq. 2)

$$V_1 = 1 \times (1+\alpha \times (d+\Delta))^{-\beta/\alpha}$$
$$V_2 = 2 \times (1+\alpha \times d)^{-\beta/\alpha}$$
$$\text{since } V_1 = V_2 \text{ We have:} (1+\alpha \times (d+\Delta))^{-\beta/\alpha} = 2 \times (1+\alpha \times d)^{-\beta/\alpha}$$
$$\Delta = ((k-1) \times (1+\alpha d))/\alpha$$
$$k = \sqrt[-\beta/\alpha]{2}$$

In other words, the hyperbolic discounting function allows   be a linear function of distance (*d*). However, introducing a new slope parameter when   is apparently constant is not parsimonious.

5. The relationship between performance and discounting factor

In this game, a player's performance is defined as the total score during the period when the red opponent, which can reduce the player's score, is absent (Table A2). Subjects playing the game accumulated an average of 354 ± 78 (standard deviation) points during the blue condition, and, d_0, the only parameter in our model of V, was consistent among subjects (21.2 ± 2.4), even though they were neither trained nor monetarily motivated to be consistent. Fig. A3 shows that the performance was not correlated with d_0 for our data.
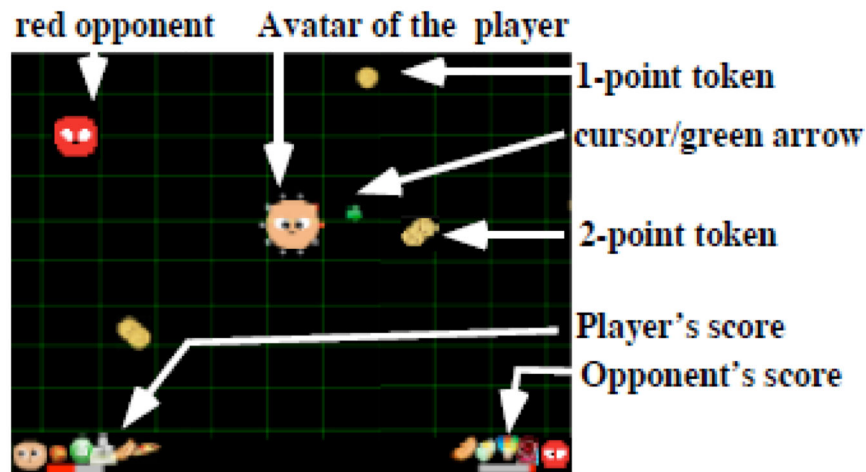

Fig. A1.
Game Scene (also see video movieS1.avi). The avatar of the player is always shown at the center (first person point of view). The green arrow on the screen indicates the direction of motion change, controlled by the player with a MRI compatible trackball. The cumulative scores of both players are presented at the bottom. For example, the avatar of the player has 55 points, which is presented by 5 bonus icons and the red color filling 5/10 of the score bar.
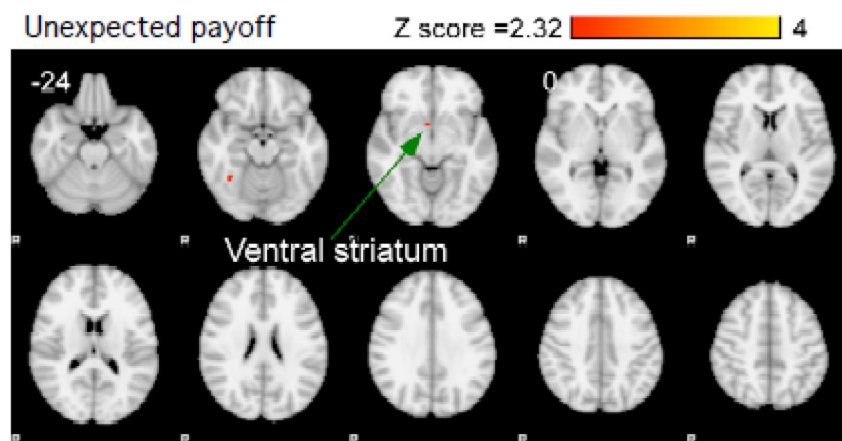
Fig. A2.

Brain activity related to the unexpected value regressor (uncorrected for multiple comparisons). The green arrow identifies a small cluster of activity in the ventral striatum.
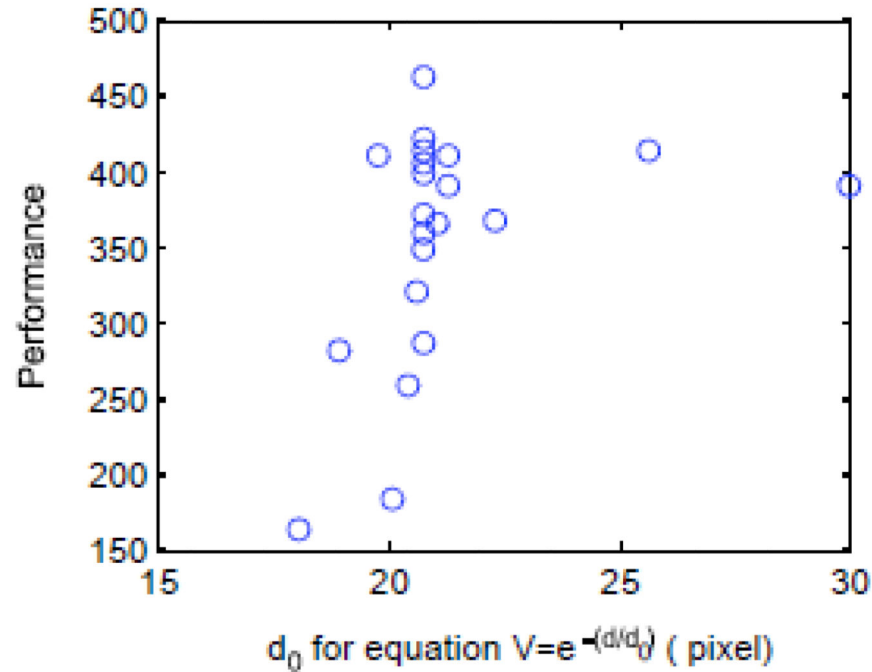

Fig. A3.

The total score over 4 runs, during blue conditions, vs. the discounting constant, the d_0 parameter for the model of the expected utility

$$V = e^{-d/d_0}.$$

References

Basar T, Olsder G. Dynamic Noncooperative Game Theory. Society for Industrial and Applied Math (2nd). 1999

Bernoulli D. Exposition of a New Theory on the Measurement of Risk. Econometrica. 1738; 22:22–36.

Glover GH, Thomason ME. Improved combination of spiral-in/out images for BOLD fMRI. Magn Reson Med. 2004; 51:863–868. [PubMed: 15065263]

Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. J Neurosci. 2008; 28:5623–5630. [PubMed: 18509023]

Isaacs, R. the theory of differential games. Wiley; New York,NY: 1965.

Kable JW, Glimcher PW. The neural correlates of subjective value during intertemporal choice. Nat Neurosci. 2007; 10:1625–1633. [PubMed: 17982449]

Loewenstein G, Prelec D. Anomalies in Intertemporal Choice - Evidence and an Interpretation. Quarterly Journal of Economics. 1992; 107:573–597.

Mathiak KA, Klasen M, Weber R, Ackermann H, Shergill SS, Mathiak K. Reward system and temporal pole contributions to affective evaluation during a first person shooter video game. BMC Neurosci. 2011; 12:66. [PubMed: 21749711]

McClure SM, Laibson DI, Loewenstein G, Cohen JD. Separate neural systems value immediate and delayed monetary rewards. Science. 2004; 306:503–507. [PubMed: 15486304]

McNamara FN, Clifford JJ, Tighe O, Kinsella A, Drago J, Fuchs S, Croke DT, Waddington JL. Phenotypic, ethologically based resolution of spontaneous and D(2)-like vs D(1)-like agonist-induced behavioural topography in mice with congenic D(3) dopamine receptor "knockout". Synapse. 2002; 46:19–31. [PubMed: 12211095]

Moncayo J, de Freitas GR, Bogousslavsky J, Altieri M, van Melle G. Do transient ischemic attacks have a neuroprotective effect? Neurology. 2000; 54:2089–2094. [PubMed: 10851368]

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

O'Doherty JP, Hampton A, Kim H. Model-based fMRI and its application to reward learning and decision making. Ann N Y Acad Sci. 2007; 1104:35–53. [PubMed: 17416921]

Opris I, Hampson RE, Deadwyler SA. The encoding of cocaine vs. natural rewards in the striatum of nonhuman primates: categories with different activations. Neuroscience. 2009; 163:40–54. [PubMed: 19501630]

Pine A, Seymour B, Roiser JP, Bossaerts P, Friston KJ, Curran HV, Dolan RJ. Encoding of Marginal Utility across Time in the Human Brain. Journal of neuroscience. 2009; 29:9575–9581. [PubMed: 19641120]

Richards JB, Mitchell SH, DeWit H, Seiden LS. Determination of discount functions in rats with an adjusting-amount procedure. Journal of the Experimental Analysis of Behavior. 1997; 67:353–366. [PubMed: 9163939]

Roesch MR, Olson CR. Neuronal activity dependent on anticipated and elapsed delay in macaque prefrontal cortex, frontal and supplementary eye fields, and premotor cortex. Journal of Neurophysiology. 2005; 94:1469–1497. [PubMed: 15817652]

Samuelson P. A note on measurement of utility. Review of Economic Studies. 1937; 4:155–161.

Sarinopoulos I, Grupe DW, Mackiewicz KL, Herrington JD, Lor M, Steege EE, Nitschke JB. Uncertainty during anticipation modulates neural responses to aversion in human insula and amygdala. Cereb Cortex. 2010; 20:929–940. [PubMed: 19679543]

Scholten M, Read D. The Psychology of Intertemporal Tradeoffs. Psychological Review. 2010; 117:925–944. [PubMed: 20658858]

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science. 1997; 275:1593–1599. [PubMed: 9054347]

Schweighofer N, Shishida K, Han CE, Okamoto Y, Tanaka SC, Yamawaki S, Doya K. Humans can adopt optimal discounting strategy under real-time constraints. Plos Computational Biology. 2006; 2:1349–1356.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang YY, De Stefano N, Brady JM, Matthews PM. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 2004; 23:S208–S219. [PubMed: 15501092]

Strotz RH. Myopia and Inconsistency in Dynamic Utility Maximization. Review of Economic Studies. 1956; 23:165–180.

Sutton RS, Barto AG. A temporal-difference model of classical conditioning. Proceedings of the Ninth Annual Conference of the Cognitive Science Society. 1987:355–378.

Watanabe M. Prefrontal Unit-Activity during Delayed Conditional Go/No-Go Discrimination in the Monkey .2. Relation to Go and No-Go Responses. Brain Research. 1986; 382:15–27. [PubMed: 3768673]

Wunderlich K, Rangel A, O'Doherty JP. Neural computations underlying action-based decision making in the human brain. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:17199–17204. [PubMed: 19805082]

Abbreviations

V	Expected utility

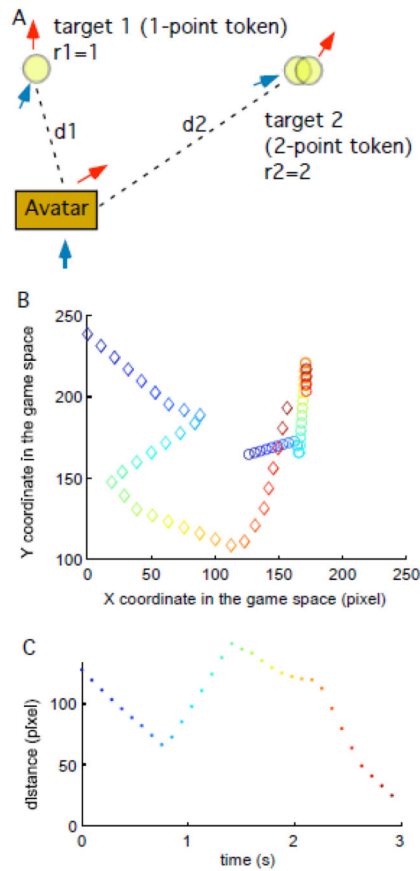| $\partial V/\partial t$ | the temporal derivative of the expected utility |
SMA	Supplementary motor area
MFC	Medial frontal cortex
ACC, PCC	anterior and posteror cingulate
VP	ventral pallidum
CPU	caudate-putaman
NAC	nucleus accumbens
CN	caudate nucleus
SPL	Superior parietal lobule

Fig. 1.
A. An example of the target-choice task. Based on the values, r1 and r2, and the distances, d1 and d2, the player makes a choice, which, in this case, is to pursue target 2. The blue arrows indicate the directions of movement prior to the event. The red arrows indicate the new directions of the movement as the result of the actions of the characters. **B.** A pursuit process. The diamond and the circles represent the location of the avatar and the target, respectively. The time is coded in color, where the blue and red colors indicate the starting and the ending time points, respectively. Each dot represents a 100-ms period. The distance between the avatar and the target is defined with dots of identical color, i.e. at the same time. Note that, due to the random perturbation of the game program, not all actions appear optimal. **C.** The game state is represented as the distance between the avatar and the target. Each dot represents the distance between the diamond and the circle shown in Fig. 1B with identical color, or at the same time.
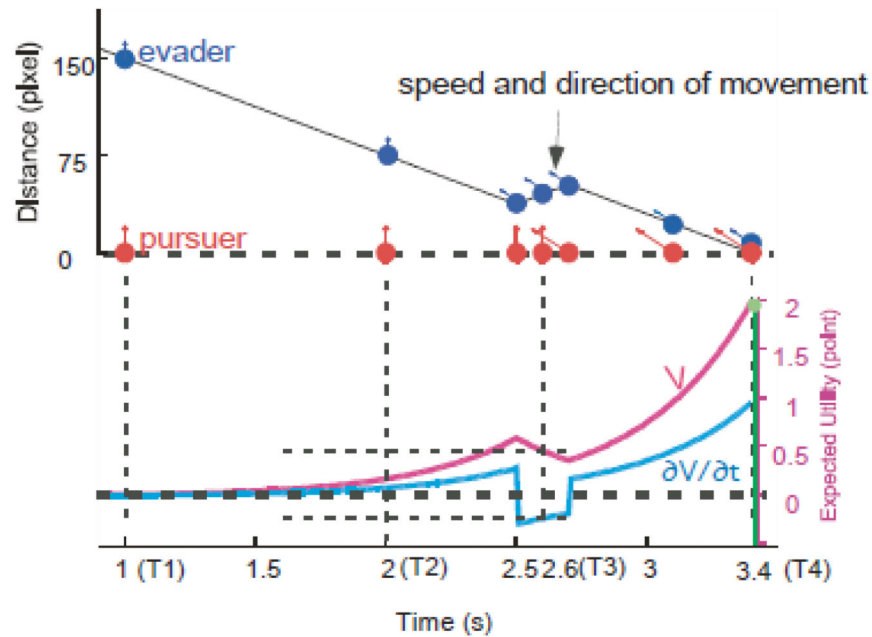
Fig. 2.

Schematic of a pursuit epoch. The solid black line in the upper panel indicates the game state, similar to Fig. 1C plotted against time. Both the relative location and the direction of movement of the avatar (red) and the target, which is a 2-point token, (blue) are shown at certain time points. The speed and the direction of movement at a time point are represented as the length and the angle, respectively, of the arrow attached to the dot. Between 1 and 2.5 second, the distance shortens because the avatar is faster. At 2.5 second, the target changed its direction but the avatar did not. As a consequence, the separation distance became greater. The avatar corrected the direction at 2.7 second. Finally, the avatar captured the target at 3.4 second and received the reward of 2 units of utility (points). In the lower panel, V (magenta) and $\partial V/\partial t$ (cyan) represent the expected utility and its temporal derivative calculated according the game state presented in the upper panel. The function of V is determined as an exponential discounting function, based on the behavioral target-choice data. The capture event is represented as an impulse function (green bar) at T4 with the magnitude of the value of the target.
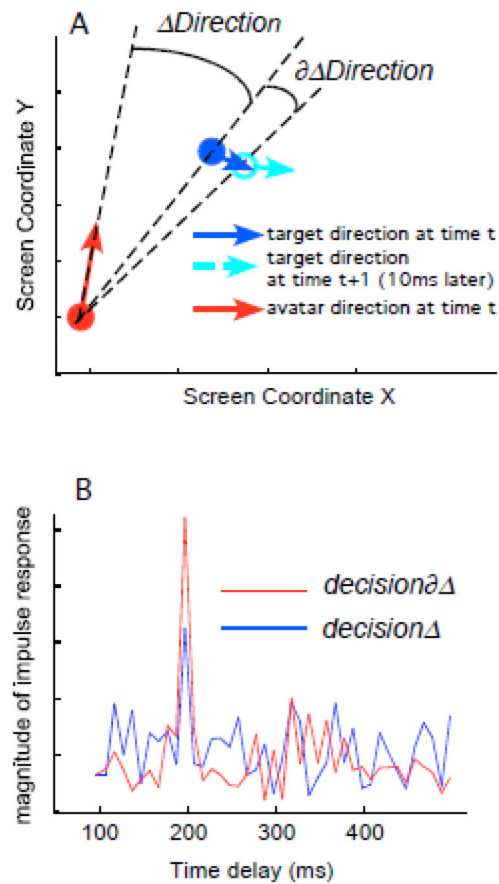
Fig. 3.

A. The positions (circles) and the directions (arrows) of the avatar (red) and the target (blue) at two time points 10 ms apart. (The direction and position changes are exaggerated for illustration purposes.) The angle arcs are graphic presentations of Δ*Direction* and $\partial\Delta$*Direction, as labeled.* **B.** The decision functions *decision*$\partial\Delta$ (red) and *decision*Δ (blue) during the last 1.5 second of the pursuit process. The peak of the decision function at 200ms indicates that, given an input time series, the action is similar to the input but delayed for 200ms. The greater magnitude of the decision function related to $\partial\Delta$*Direction* indicates the significance of the temporal derivative.
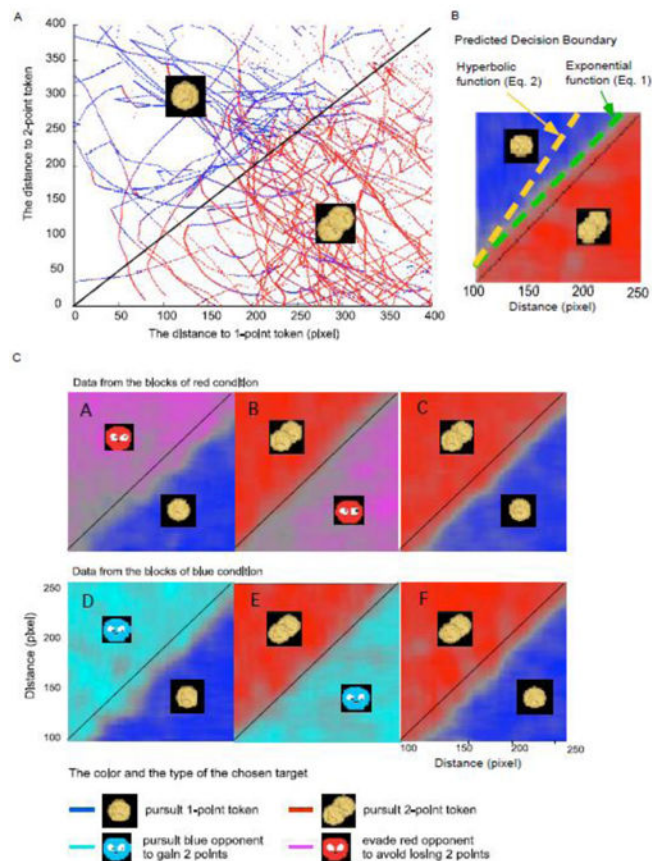
Fig. 4.
A. Distance vs. Distance Scatter-plot showing an example of a subject's pursuit choices as a function of distance and the value of the targets during an 8-minute run. Each dot represents a 10ms time point and the distance in pixels to the 1-point token (X axis) and the distance to the 2-point token (Y axis). Red and blue indicate that the player is pursuing the 2-point token and the 1-point token, respectively. A line represents a series of consecutive measures. If a red line reaches the X-axis, the 2-point token is captured and if a blue line reaches the Y-axis, the 1-point token is captured, i.e. the distance becomes zero. The two tokens shown inside either the upper or the lower triangle indicate the target of choice for most of the time. **B.** The averaged and smoothed pursuit choices for 21 subjects and 4 runs in the distance range of between 100 to 250 pixels. Saturated colors indicate that the icon shown in the region was chosen most of the time, while gray indicates the decision boundary, where the probability for choosing either target is close to 0.5. The theoretical predictions by both the exponential discounting function (green dashed line) and the hyperbolic function (orange dashed line) are also presented. **C.** The average *pursuit-choices* across 21 subjects and 4 runs. See Table A1 in the Appendix for the payoff tables related to these decisions. In short, panels A,D,C and F involve a choice between two points and one points, while panels B and E involve a choice between two options of two points. The distance ranges from 100 to 250 pixels, where the data quality is high.
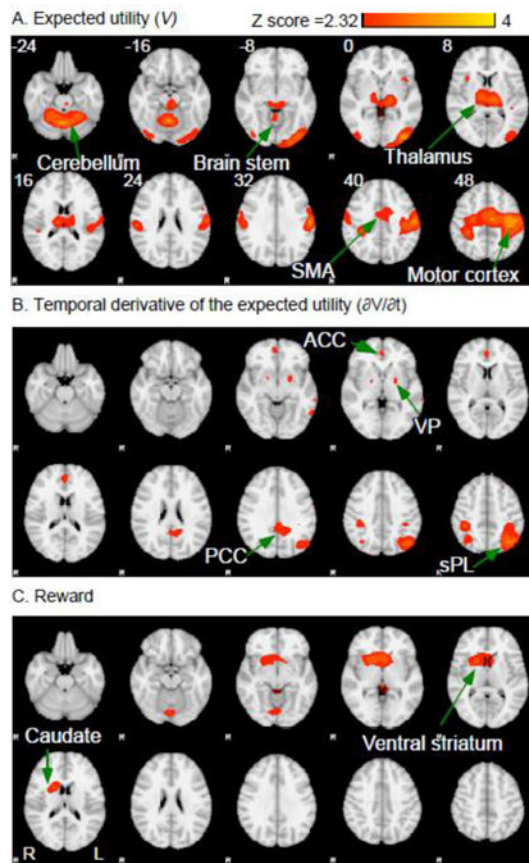
Fig. 5.
Areas for which brain activity is correlated with regressors V (**A**), $\partial V/\partial t$ (**B**) and Capture event (**C**). Red/yellow indicates that the BOLD signal is positively correlated with the regressors and the Z scores range from 2.33 to 4. The images are presented following radiological convention, where the right brain is presented as image left. The abbreviations in the figure include anterior Cingulate Gyrus (ACC), posterior Cingulate Gyrus (PCC), ventral Pallidum (VP) , supplementary motor (SMA) and superior Parietal Lobule (SPL).

Table 1

Z-scores, MNI Coordinates and Brain Areas of the Local Maxima. Brain areas are identified with the Harvard-Oxford cortical structure atlas. The significance of the activity clusters were corrected for multiple-comparisons at p value < 0.05.

A. V

Brain area for the local maxima	Hemi				score	Z
Precentral Gyrus (PRECG)	L	36	14	4	.39	4
Precentral Gyrus (PRECG)	R	4	12	0	.3	4
Thalamus (THAL)	L	14	20		.9	3
Thalamus (THAL)	R	2	16	0	.45	3
inferior lateral Occipital Cortex (ILOC)	L	38	88		.09	4
inferior lateral Occipital Cortex (ILOC)	R	2	86		.89	2
Postcentral Gyrus (POSTCG)	L	40	24	0	.32	4
Supplementary motor area (SMA)	L	4	2	6	.98	3
Brain-Stem	L	6	30	14	.18	3
	R					

B. V/ t

Brain area for the local maxima	Hemi				score	Z
superior lateral Occipital Cortex (SLOC)	L	42	68	6	.84	3
superior lateral Occipital Cortex (SLOC)	R	0	56	6	.25	3

A. V

Brain area for the local maxima	Hemi				score	Z
posterior Cingulate Gyrus (PCC)	L	14	48	8	.79	2
posterior Cingulate Gyrus (PCC)	R		44	0	.74	2
ventral Pallidum (VP)	L	22	6		.69	2
posterior Middle Temporal Gyrus (PMTG)	R	8	38	8	.3	3
superior Parietal Lobule (sPL)	L	36	44	2	.21	3
anterior Cingulate Gyrus (ACC)	R		4	4	.84	2
medial Frontal Cortex (MFC)	R		2	10	.62	2
Frontal Inferior Orbital Gyrus (IFG)	L	54	8	4	.46	2
Precentral Gyrus (PRECG)	L	62	2	4	.4	2
Postcentral Gyrus (POSTCG)	L	64	10	0	.34	2
superior Frontal Gyrus (SFG)	L	26	8	6	.38	2
Frontal Superior Medial Gyrus	R		0	2	.37	2

C. Capture event

Brain area for the local maxima	Hemi				score	Z
Subcallosal Cortex (SCC)	R		0		.56	3

A. V

Brain area for the local maxima	Hemi				Z score	Z
Putamen (CPU)	R		2	2	.46	3
Caudate Nucleus (CN)	R	0	2		.39	3
Lingual Gyrus (LING)	R		78	14	.23	3
Brain-Stem	L		40	4	.89	2
	R					