Self-Controlled Choice Arises from Dynamic Prefrontal Signals That Enable Future Anticipation

Daiki Tanaka,¹ Ryuta Aoki,² Shinsuke Suzuki,^{3,4} Masaki Takeda,² Kiyoshi Nakahara,² and Koji Jimura^{1,2}

¹Department of Biosciences and Informatics, Keio University, Yokohama 223-0061, Japan, ²Research Center for Brain Communication, Kochi University of Technology, Kami 782-8502, Japan, ³Brain, Mind & Markets Laboratory, Department of Finance, Faculty of Business and Economics, The University of Melbourne, Victoria 3010, Australia, and ⁴Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai 980-8575, Japan

Self-control allows humans the patience necessary to maximize reward attainment in the future. Yet it remains elusive when and how the preference to self-controlled choice is formed. We measured brain activity while female and male humans performed an intertemporal choice task in which they first received delayed real liquid rewards (forced-choice trial), and then made a choice between the reward options based on the experiences (free-choice trial). We found that, while subjects were awaiting an upcoming reward in the forced-choice trial, the anterior prefrontal cortex (aPFC) tracked a dynamic signal reflecting the pleasure of anticipating the future reward. Importantly, this prefrontal signal was specifically observed in self-controlled individuals, and moreover, interregional negative coupling between the prefrontal region and the ventral striatum (VS) became stronger in those individuals. During consumption of the liquid rewards, reduced ventral striatal activity predicted self-controlled choices in the subsequent free-choice trials. These results suggest that a well-coordinated prefrontal-striatal mechanism during the reward experience shapes preferences regarding the future self-controlled choice.

Key words: delay discounting; intertemporal choice; prefrontal cortex; reward; ventral striatum

Significance Statement

Anticipating future desirable events is a critical mental function that guides self-controlled behavior in humans. When and how are the self-controlled choices formed in the brain? We monitored brain activity while humans awaited a real liquid reward that became available in tens of seconds. We found that the frontal polar cortex tracked temporally evolving signals reflecting the pleasure of anticipating the future reward, which was enhanced in self-controlled individuals. Our results highlight the contribution of the fronto-polar cortex to the formation of self-controlled preferences, and further suggest that future prospect in the prefrontal cortex (PFC) plays an important role in shaping future choice behavior.

Introduction

A remarkable signature of human disposition is the anticipation of unknown events that may occur in the remote future. The anticipation of the future underlies self-controlled behavior, characterized as preference for a greater amount of rewards in the future instead of a smaller amounts of reward available sooner (Berns et al., 2007; Fig. 1A). Such behavioral situations are well illustrated by intertemporal choice, in which a choice is

made between alternative reward options varying both in their magnitude and time of delivery (Keeny and Raiffa, 1993; Green and Myerson, 2004).

One reliable finding of intertemporal choice is that behavioral agents discount the value of rewards that are delayed in time (Mischel et al., 1989; Frederic et al., 2002; Ainslie, 2005). This phenomenon, called delay discounting, is thought to be related to self-controlled and impulsive choice behavior (Madden and Bickel, 2009; Peters and Büchel, 2011). In particular, individuals with strong self-control show smaller delay discounting (Kirby et al., 1999; McClure et al., 2004, 2007; Luo et al., 2009; Figner et al., 2010; Kable and Glimcher, 2007). It is natural to postulate that preference for self-controlled or impulsive choice is dependent on past experiences of reward attainment and anticipation of unexperienced future outcomes (Loewenstein, 1987; Rachlin, 2004; Fig. 1A). However, it still remains unclear when and how choice preference is formed through the experience of rewarding and anticipating events.

Interestingly, non-human animals (e.g., rodent, birds and primates) also show delay discounting in intertemporal choice

Received July 1, 2020; revised Nov. 2, 2020; accepted Nov. 4, 2020.

Author contributions: D.T. and K.J. designed research; D.T., R.A., K.N., and K.J. performed research; D.T., S.S., and K.J. analyzed data; D.T., R.A., S.S., M.T., K.N., and K.J. wrote the paper.

This work was supported by Grants-in-Aid for Scientific Research (KAKENHI) 26350986, 26120711, 17K01989, 17H05957, and 19H04914 (to K.J.); 17H00891 (to K.N.); and 20H00521, 18H04953, 18H04953, 18H05140, and 17K07062 (to M.T.); the Uehara Memorial Foundation (K.J. and M.T.); and the Takeda Science Foundation (K.J. and M.T.). We thank Dr. Kenji Miyamoto and Dr. Norifumi Kawakami for their technical supports. We thank Maoko Yamanaka for administrative assistance.

The authors declare no competing financial interests.

Correspondence should be addressed to Koji Jimura at jimura@bio.keio.ac.jp.

<https://doi.org/10.1523/JNEUROSCI.1702-20.2020>

Copyright © 2020 the authors

Figure 1. Experimental design. **A**, Intertemporal choice situation. Behavioral agent makes a choice between greater amount of rewards (several kinds of fruits) available in the future (next week), or smaller amount of rewards (one banana) available now. Self-controlled choice is characterized as a preference toward the larger later option (left). The choice is made based on past experience of rewarding events (e.g., consuming fruits) and anticipation of future events (right). **B**, Behavioral task. Two pictures were presented on the screen: one picture indicated a larger amount of a delayed reward and the other picture indicated a smaller amount of an immediate reward. **C**, Task procedure. Participants directly experienced two rewards in forced choice trials, and then freely made a choice between these options based on past experiences from the forced choice trials. These three trials (a block) were performed in triplicate. Note that the picture indicating the reward was unique for each trial block.

(Richards et al., 1997; Freeman et al., 2009; Vanderveldt et al., 2016; Carter and Redish, 2016), although several discrepancies exist between human and non-human animal studies. Human studies have suggested that the ventromedial prefrontal cortex (PFC) represents the value of the delayed reward (Kable and Glimcher, 2007; Ballard and Knutson, 2009; Sellitto et al., 2010), while non-human studies have shown that the value representation is reflected in neuronal activity in the ventral striatum (VS; Cai et al., 2011; Stott and Redish, 2014). On the other hand, the human VS is associated with behavioral impulsivity (Kable and Glimcher, 2007; Jimura et al., 2013), and neuronal activity in the non-human PFC is independent of the value representation (Roesch et al., 2006; Cai et al., 2011; Stott and Redish, 2014). One possible reason for the discrepancies of the prefrontal-striatum involvements in intertemporal choice is the differences in behavioral paradigms.

Behaviorally, in non-human animal studies, animals directly experienced a delay and were then given consumable primary rewards (e.g., water and food pellets) during an experimental session. Then, they freely made a choice between delayed and immediate reward options based on the direct experience of the rewards. In this procedure, because animals first learned the relationship between the duration of the delay and consumption of the reward through their experiences, it is possible to assume that choice preference is formed during the direct experience (Fig. 1A). This is contrasted with standard human experiments assuming that choice preference has been formed before experimental sessions and is stable through the entire experiment.

Common human paradigms of intertemporal choice have used a secondary reward such as money, and choice options have been provided in hypothetical situations (Green and Myerson, 2004). Those paradigms have limited neuroimaging studies to examine brain mechanisms only when a choice was made. In some studies, however, human participants directly experienced a delayed outcome during experimental sessions, which allowed the examination of temporal characteristics of brain activity while the participants were anticipating future outcomes (Berns et al., 2006; Jimura et al., 2009, 2011, 2013; McGuire and Kable, 2015; Iigaya et al., 2020). Importantly, these studies used economic modeling to examine the dynamics of delay period activity, in which the value of the future reward was temporally evolved as the reward outcome approaches (Jimura et al., 2013; McGuire and Kable, 2015; Iigaya et al., 2020). Nonetheless, unlike non-human studies, participants did not learn the association between delay and reward amount during experimental sessions. Thus, discrepancies still exist between human and non-human animal paradigms in terms of

the reward type and choice preference formation.

While anticipating a delayed reward during intertemporal choice behavior, the VS shows dynamically modulated time-discounted value of the reward (Cai et al., 2011; Jimura et al., 2013), reflecting behavioral impulsivity (weaker self-control; Hariri et al., 2006; Jimura et al., 2013), whereas the anterior PFC (aPFC) and orbitofrontal cortex show prolonged activity (Roesch et al., 2006; Jimura et al., 2013), reflecting strong self-control (Shamosh et al., 2008; Jimura et al., 2013). These temporally modulated value-related activations and interregional functional coupling (Diekhof and Gruber, 2010) collectively suggest a dynamic prefrontal-striatal mechanism during anticipation of a delayed reward, which reflects the degree of self-control.

The current study aimed to examine when and how self-controlled choice preference was formed from direct experience of

delayed rewards and to elucidate underlying neural mechanisms. To address this issue, we applied non-human experimental procedures of intertemporal choice to a human experiment (Fig. 1B,C). During functional MRI (fMRI) scanning, participants directly experienced delayed and immediate rewards without quantitative verbal information (forced choice trials). Then, they freely made a choice between the two options based on the past two trial experiences (free-choice trial). These three trials (a block) were performed in triplicate in an experimental session. Imaging analysis focused on the temporal dynamics of brain activity during the experience of a delayed reward based on economic and learning model approaches. Individual differences in activity dynamics were then explored to identify brain regions where activity dynamics predicted subsequent self-controlled choices.

Materials and Methods

Participants

Human participants ($N=34$; age range, 18–22 years; 12 female) were right-handed and had no history of psychiatric or neurologic disorders. Written informed consent was obtained from all participants. All experimental procedures were approved by the institutional review boards of Keio University and Kochi University of Technology. Participants were instructed not to drink any liquid for 4 h before the experiment, and received 2000 yen for participation.

Reward

The current study used commercially available drinks as a reward that could be consumed directly. Before the experiment, participants were asked to choose one favorite drink that would serve as the reward from a list consisting of apple, orange, grape, lychee, and mixed fruit juices, sports drink, probiotic drinks (plain and muscat grape flavor), and water.

Apparatus

E-Prime programs (Psychology Software Tools) controlled the task as well as the delivery of liquid rewards via a syringe pump (SP210iw; World Precision Instruments). Liquids from two 60-ml plastic syringes mounted on the pump were merged into one tube and then delivered to the participant's mouth through a silicon tube. The flow rate of each syringe was set to 0.75 ml/s, and thus the reward flowed continuously at a flow rate of 1.5 ml/s. Participants were able to control the liquid flow. Reward delivery continued as long as they pressed down a button on a button box that they held in their right hand; if they released the button, delivery paused and then resumed when they pressed the button again.

Behavioral procedures

During fMRI scanning, human participants performed an intertemporal decision-making task of real liquid rewards delayed by seconds. At the beginning of each trial, two pictures were presented (Fig. 1B). One picture indicated a delayed reward (12 ml; 10, 30, or 60 s), and the other indicated an immediate reward (1.5–10.5 ml). Importantly, participants were unfamiliar with the amount of the rewards, the duration of the delay, and the relationships between the pictures and rewards, like in non-human studies.

One trial block consisted of three types of trials (Fig. 1C, top). In the first trial, participants were presented with a visual message to choose a picture, and forced to choose the picture indicating the delayed reward (forced choice trial). Choice stimuli were presented until the participants' response. Participants made a response by pressing a button (left or right) with their right thumb. When participants pressed the corresponding button, the delay period started, and the other picture was disappeared. After the delay finished, a visual message indicating that the reward was ready was presented, and participants consumed the liquid reward. After they consumed all of the liquid reward, the choice picture was replaced by a fixation cross. Then, two identical pictures were presented again, and participants were asked which picture they chose in the previous choice. They then received feedback for their choice depending on correct or incorrect response (probe test; Fig. 1C,

bottom). This probe trial was intended to ensure that participants learned the relationships between picture choice and rewards.

The next trial started after intertrial interval. Two identical pictures were presented again, and participants were presented with a visual message to choose another picture, and forced to choose the other picture indicating immediate reward (forced choice trial). Immediately after participants pressed the corresponding button, the picture indicating the delayed reward disappeared and a message indicating that drink was ready was presented. Participants then consumed the liquid reward, followed by another probe trial with the immediate reward picture.

The forced choice trial of the delayed reward was always followed by the forced choice trial of the immediate reward. Because the primary focus of the current experiment was the brain activity dynamics during delay period of the forced choice trial, the current task was designed such that the delay period was unaffected by the experience of an immediate reward. It should be noted that the amount of the immediate reward was adjusted to estimate subjective value (SV) of the delayed reward (see this section below). Thus, if the immediate reward trial had preceded the delayed reward trial, the experience of the delay would be affected by the variable experiences of the preceding immediate reward. As such, by presenting the delay period of the forced choice trials first in each block, we ensured that participants experienced the delay without contamination from the immediate reward in the block.

One potential confounding factor of this design is that the experience of the immediate reward might be affected by the preceding experience of the delayed reward. Use of a constant amount of the delayed reward (12 ml) and delay duration (10, 30, or 60 s; see also below) was intended to minimize this confounding factor, however.

After participants experienced both of delayed and immediate rewards, two identical pictures were presented again on the screen. In the third trial, participants freely chose options as they preferred (free-choice trial). Depending on their choice, participants directly received the delayed or immediate reward, where the reward amount and delay were identical to those in the prior two trials. After reward consumption, participants were asked whether they chose intended choice option (choice confirmation; Fig. 1C, bottom).

In the forced choice trials, free-choice trials, and probe tests, the positions of the pictures (left or right) indicating the delayed and immediate rewards were randomized within a trial block to prevent participants from learning the association of the pictures and rewards based on the placements.

As stated above, one trial block consisted of two forced choice trials and one free-choice trial, and participants performed a total nine blocks during fMRI scanning. Before starting a block, a visual message was presented to inform participants of the start of the next block, and they were required to press either button.

The duration of delay for the delayed reward was either of 10, 30, or 60 s, and the amount of the delayed reward was kept at a constant 12 ml. Each delay condition was presented once in one scanning run. Trial blocks with different delay conditions were presented pseudorandomly.

The amount of immediate reward was 6 ml for the first block of each delay condition. In order to calculate the SV of the delayed reward, the amount of the immediate reward from the second block was adjusted according to the preceding choice (Jimura et al., 2009, 2011, 2013). If participants had chosen the immediate reward on the preceding free-choice trial, then the amount of the immediate reward was decreased by half; if participants had chosen the delayed reward on the preceding trial, then the amount of the immediate reward was increased by half. The SV for the delayed reward was estimated to be equal to 0.75 ml more or less than the amount of the immediate reward available on the last trial, depending on whether the delayed reward or immediate reward had been chosen in that trial.

Before fMRI scanning, participants received instructions for the task using a computer display. Participants were told that the two pictures indicated a delayed reward or immediate reward and asked to remember the delay duration and reward amount through the experience of forced choice trials. In order to familiarize participants with task, one block was performed as a practice block (delayed reward: 20 s, 9 ml; immediate

reward: 3 ml). Participants were not informed that the practice block was a practice.

We used 18 unique pictures (six shapes $\times$ three colors) for the intertemporal choice task, and one picture was used in only one block. Participants were told that pictures used in one task block are unrelated to other task blocks. Choice pictures used in the practice block were not used in scanning blocks.

Assessment of self-control

Individual differences in delay discounting were quantified by calculating the area under the curve (AuC) of individuals' discounting plots (Myerson et al., 2001; Sellitto et al., 2010; Jimura et al., 2011, 2013). The AuC represents the area under the observed SVs. The AuC was calculated as the sum of the areas of three trapezoids under the SV points (10, 30, and 60 s) from 0 to 10, 10 to 30, and 30 to 60 s, divided by the longest delay duration (60 s). Specifically,

$$AuC = \frac{\frac{(1 + SV_{10}) \times 10}{2} + \frac{(SV_{10} + SV_{30}) \times 20}{2} + \frac{(SV_{30} + SV_{60}) \times 30}{2}}{60}, \quad (1)$$

where SV_{10} , SV_{30} , and SV_{60} denote SVs of the delayed reward standardized by the reward amount (12 ml), on 10, 30, and 60 delay trials, respectively. It has been argued that the AuC is a valid measure of delay discounting to use for individual difference analyses, because it is theoretically neutral and also psychometrically reliable (Myerson et al., 2001).

Imaging procedure

fMRI scanning was conducted on a whole-body 3T MRI system (Siemens Verio). Functional images were acquired using multiband accelerated gradient-echo echo-planar imaging [repetition time (TR) = 800 ms; echo time (TE) = 30 ms; flip angle (FA) = 45°; slice thickness, 2 mm; in-plane resolution, 3×3 mm; multiband factor (MBF) = 8; 80 slices]. Whole-brain scanning with high temporal resolution allowed us to perform whole-brain exploratory analyses of temporal dynamics during the delay period with sufficient scanning frames. Each run involved 1088 volume acquisitions (14.5 min), and three functional runs were performed (total 3264 volumes). The initial 10 volumes of each run were excluded from imaging analysis to take into account the equilibrium of longitudinal magnetization. High-resolution anatomic images were acquired using an MP-RAGE T1-weighted sequence (TR = 2500 ms; TE = 4.32 ms; FA: 8°; 192 slices; slice thickness, 1 mm; in-plane resolution, 0.9×0.9 mm²).

Imaging analysis procedures

Preprocessing

Functional images were preprocessed using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>). All functional images were first temporally aligned across the brain volume, corrected for movement using a rigid-body rotation and translation correction, and then registered to the participant's anatomic images to correct for movement between the anatomic and function scans. Participants' anatomic images were transformed into standardized MNI templates. The functional images were then registered to the reference brain using the alignment parameters derived from the anatomic scans. The data were then resampled into 2-mm isotropic voxels, and spatially smoothed with a 6-mm full-width at half-maximum Gaussian kernel.

In order to minimize motion-derived artifacts because of consumption of liquid rewards, functional images were further preprocessed by general linear model (GLM) estimations with motion parameters and MRI signal time courses (CSF, white matter, and whole brain), and their derivatives and quadratics as nuisance regressors (Ciric et al., 2017; Keerativittayayut et al., 2018). The residual of the nuisance GLM was used for standard GLM estimations to extract task event-related brain activity described below, based on *fsl_regfilt* implemented by FSL suite (<http://www.fmrib.ox.ac.uk/fsl/>; version 5.0.8).

Head motions were greater along the z-translation and x-rotation axes during liquid consumption than the button press, but the mean magnitude of absolute motion was <0.458 mm for translation and 6.33×10^{-3}

radian for rotation, which is consistent with our prior study (Jimura et al., 2013). Critical image distortions or signal drop-outs were not observed, but robust brain activity was extracted in the primary motor and gustatory cortex during liquid consumptions, which assured us that motion derived artifacts did not critically contaminate the current analyses, like our prior study (Jimura et al., 2013).

GLM

Single level analysis

A GLM approach was used to estimate trial event effects. Parameter estimates were performed by *feat* implemented in the FSL suite. Events of particular interest were delay periods of forced choice trials for the delayed reward, and of free-choice trials where the delayed reward was chosen. The current analyses focused on the temporal dynamics of brain activity during these events.

The brain activity during the delay period was represented by two value representation models. One was the anticipatory utility (AU) model, which reflected the utility of waiting for a future reward (Loewenstein, 1987; Berns et al., 2006; Jimura et al., 2013), and the other reflected the increase in the future reward value as the delay passed (Montague and Berns, 2002; Green and Myerson, 2004; Kalenscher and Pennartz, 2008; Jimura et al., 2013). AU showed a peak value at the beginning of the delay and a gradual decrease along the course of delay, while the future reward value model had inverse temporal characteristics.

One key issue to model delay period dynamics during the forced choice trials was that participants were unfamiliar with the upcoming delayed reward, because they had no experience with the reward indicated by the presented picture that was unique for each trial block (see above, Behavioral procedures). Nonetheless, it was possible to expect when a future reward would become available based on past trials they experienced in prior blocks in which distinct pictures were used to indicate the rewards. Thus, the current modeling assumed that, before the trial, participants had an expectation of the duration of the delay, and the expectation was updated through the experience of forced trials for the delayed reward.

The current analysis modeled the expectation of duration and its updates based on Bayesian inference learning approach (Behrens et al., 2007). We assumed that the expectation was expressed by a probability density distribution of the reward outcome, and the distribution was calculated as a function of delay time using a γ function. Then, the distribution was updated every time after participants completed forced choice trial.

Bayesian inference was performed as:

$$P(\theta|D) \propto P(\theta)P(D|\theta), \quad (2)$$

where D represents delay duration, and θ is a parameter that determines probability density. Then probability density function (PDF) was calculated as

$$PDF = P(D) = \int P(D|\theta)P(\theta)d\theta, \quad (3)$$

where $P(D|\theta)$ takes a γ distribution. The average of the initial distribution was 20 s (i.e., the delay duration in the practice trial), and the distribution became wider as the forced choice trials for the delayed rewards were experienced (Fig. 2A).

In order to model AU dynamics during the delay period, the PDF was first integrated from the start of the delay to time t , defined as cumulative probability (CP):

$$CP(t) = \int_0^t PDF(t)dt. \quad (4)$$

The idea of this integration is that, as the delay period elapsed, participants' expectation of reward outcome continued to increase, although

Figure 2. Modeling value representation dynamics during delay period of the forced choice trial. **A**, Probability density distribution for expected delay was estimated on based on Bayesian inference and updated through the experiences of forced choice trials for delayed reward. **B**, CP was then computed as an integral of the PDF from the start of the delay to time t , and the dynamics of AU was modeled as an inverse of the CP (top). Upcoming unexperienced reward value was modeled as the CP (bottom). The color of the line indicates trial experiences of the delayed reward as shown in the color bar at the bottom. **C**, AU and upcoming unexperienced reward value models were convolved with the canonical HRF.

they did not exactly know when the reward became available. The $CP(t)$ reflects an estimation of the probability that the reward is likely to become available, like hazard rates (Janssen and Shadlen, 2005; McGuire and Kable, 2015). As time t goes to infinity, $CP(t)$ is reaching to the upper limit:

$$\lim_{t \rightarrow \infty} CP(t) = 1. \quad (5)$$

Then, AU was defined as:

$$AU = 1 - CP(t) (0 < t < D),$$

$$AU = 0 (t > D) \quad (6)$$

where D denotes duration of the delay. Immediately after the delay finished and the reward became available to drink, the value of AU was set to 0. The AU shows a temporal dynamics that peaks at the beginning of the delay and is monotonically decreased as the delay period elapses (Fig. 2B, top), like prior modeling (Jimura et al., 2013).

$CP(t)$ is also associated with a value component of a future reward that was not yet experienced, but would become available eventually, like prior modeling (Janssen and Shadlen, 2005; Berns et al., 2006; Jimura et al., 2013; McGuire and Kable, 2015). Then the current study defined the future reward value as the upcoming unexperienced future reward (UFR) as $CP(t)$ (Fig. 2B, bottom; Janssen and Shadlen, 2005; McGuire and Kable, 2015):

$$UFR(t) = CP(t) (0 < t < D),$$

$$UFR(t) = 0 (t > D) \quad (7)$$

The value of UFR was set to 0, immediately after the delay finished and the reward became available to drink, similarly to AU value.

Then, the AU and upcoming unexperienced reward value models were calculated for each forced choice trial for each participant, and then convolved with a canonical hemodynamic response function (HRF; Fig. 2C).

By definition, in the economic models, AU and UFR are anticorrelated, but when convolved with the HRF, they produced dissociable BOLD regression functions (Fig. 2C). Specifically, the correlation between the BOLD models of AU and UFR is 0.41 ± 0.01 (mean $\pm$ SD), allowing sufficient dissociation (cf. Otten et al., 2002). Additionally, we also performed separate control GLM analyses, where only one of the two models was coded in GLM. In those control analyses, the UFR and AU effects were reasonably reproduced, confirming that multicollinearity of those regressors was not an issue, consistent with our previous study (Jimura et al., 2013).

For the free-choice trial, to model temporal dynamics of value representation during the delay period, the two components were defined based on linear functions (Berns et al., 2006), rather than the discounting curve to circumvent circular analysis (Kriegeskorte et al., 2009). The delay period started after they chose the delayed reward, and participants experienced a delayed reward identical to the one that they directly experienced immediately before they made a choice. Thus, a discounting curve may not be optimal, because the discounting curve was estimated based on choices in the free-choice trials, although discounting curve was available to model delay period dynamics in our prior study (Jimura et al., 2013), because the SVs were estimated in a separate behavioral session before a scanning session (for more details, see Discussion). The AU model showed the maximal value (1) at the beginning of the delay period and a gradual decrease the minimum value (0) at the end of the delay. We used this formulation because during the delay period of the free-choice trial participants knew when the delay finished from the experience of the forced choice trial. Specifically,

$$AU = 1 - \frac{t}{D} (0 < t < D),$$

$$AU = 0 (t > D). \quad (8)$$

The SV model showed the inverse dynamics, formulated as the following:

$$SV = 1 - AU = t/D(0 < t < D),$$

$$SV = 0(t > D). \quad (9)$$

Then, the two dynamic value components were convolved with canonical HRF.

In order to examine brain activity during liquid consumption of the forced choice trials, the trials were classified as reward (delayed or immediate) and subsequent choice in free-choice trial (chosen or unchosen), entailing a 2×2 factorial model. The chosen/unchosen factor classified the trials whether the reward option was chosen in the subsequent free-choice trials in the same block, and is orthogonal to the delayed/immediate factor. Thus, the 2×2 model consisted of (1) chosen immediate, (2) unchosen immediate, (3) chosen delayed, and (4) unchosen delayed reward trials. Then, the consumption events of the four types of forced choice trials were separately coded in GLM as a box-car epoch starting from the onset of consumption and lasting until completion of consumption, and were convolved with canonical HRF.

Other events of non-interest involved choice during forced and free-choice trials, probe tests, liquid consumption of free-choice trial, and message presentation for the next trial block. These events were coded separately, convolved with canonical HRF.

Group-level analysis

Parameter estimation maps were collected from all participants and subjected to a group-level one-sample t tests based on permutation methods (5000 permutations) implemented by *randomize* in FSL suite. Voxel clusters were identified using a voxel-wise uncorrected threshold of $p < 0.01$, and the voxel clusters were tested for a significance with a threshold of $p < 0.05$ corrected by family-wise error (FWE) rate. This non-parametric permutation procedure was validated to control false positive rate, appropriately (Eklund et al., 2016). The peaks of significant clusters were then identified and listed on tables. If multiple peaks were identified within 12 mm, the most significant peak was retained. In individual difference analyses for the forced choice trial, correlation maps were masked by the aPFC region showing an AU effect for the free-choice trial.

In order to evaluate Bayesian learning model, we compared the temporal dynamics of AU models and aPFC signals during the first and last (ninth) trials. Specifically, for each participant, we extracted time courses of BOLD models of AU and aPFC signal reflecting the initial 10-s delay epoch (the longest duration in which data were available from all participants) during the first and last trials, and then averaged them across participants. Note that the 10-s behavioral epoch corresponds approximately to a 0- to 26-s window in BOLD models and signal time courses, given the hemodynamic signal in the human brain.

For the analysis related to the free-choice trials, the number of the choice for the delayed and immediate rewards varied across participants (see Results). The effect of imbalanced trial sample was partially reduced by the group-level mean analysis treating participants as random effect, where only cross-subject error was considered. Nonetheless, because of this imbalanced trial sample, for the free-choice trials, we did not perform individual difference analyses (i.e., correlation analysis with the discounting measure), in which the number of the choice directly contaminated the results.

Functional connectivity analysis

A hierarchical mixed-effect GLM (Raudenbush and Bryk, 2002) was used to examine functional connectivity between the aPFC and VS. Because this modeling is based on GLM, a causal pathway cannot be tested between regions but rather it assumes a connectivity pattern in which the VS receives inputs from the aPFC during anticipation of a future reward (Jimura et al., 2013), as suggested by prior work (Koechlin and Hyafil, 2007; Schacter et al., 2007; Rolls et al., 2008; Glimcher, 2009; Hare et al., 2009; Peters and Büchel, 2011; Benoit et al., 2011; Daw et al., 2011).

The model included two levels, one for within-subject trial-by-trial effects, and the other for between-subject effects. The lower level within-subject effect was modeled as the following:

$$X_{VS} = \beta_{aPFC} X_{aPFC} + \beta_{const} + \varepsilon_{VS}, \quad (10)$$

where X_{VS} and X_{aPFC} indicate MRI signal during delay period in VS and aPFC, respectively, β -values represent regression coefficients, and ε_{VS} is an error term. The higher level between-subject effect was modeled as the following:

$$\beta_{aPFC} = \gamma_{AuC-aPFC} X_{AuC} + \gamma_{aPFC-const} + \varepsilon_{aPFC}, \quad (11)$$

where X_{AuC} indicates AuC values of individual participants, γ values indicate regression coefficients, and ε_{aPFC} is an error term.

Regions of interest (ROIs) were defined independently of the tested effects to avoid circular analysis. VS ROI was defined anatomically based on the Harvard-Oxford Atlas. For delay effect of the forced choice trial, aPFC ROI was based on the aPFC region showing AU effect during free-choice trial and vice versa (i.e., to test the delay period of free-choice trial, ROI was defined based on forced choice trials). More specifically, peak coordinates were identified in the aPFC region, and spheres with radii of 6 mm centered on the peaks were created.

Then in these ROIs, the fMRI signal time courses relative to the fixation baseline were extracted during the delay period for forced and free-choice trials. Scanning frames from 5 to 10 (10-s trial), from 5 to 30 (30-s trial), and from 5 to 60 (60-s trial) were extracted and then averaged across time courses for each trial. These trial-by-trial signal values were submitted to the model and all parameters were simultaneously estimated using the *lmer* procedure in R (<http://www.r-project.org/>). Statistical testing was performed using the *lmerTest* procedure.

Data and code availability

The datasets and codes supporting the current study are available from the corresponding author on reasonable request.

Results

Delay discounting of a real liquid reward without quantitative information

In the current experiment, participants first experienced delayed and immediate rewards followed by a free-choice trial where they chose one of the two options as they preferred based on past experiences of the two forced choice trials (Fig. 1C). If they chose a delayed reward, then they received the same delay and liquid reward. If they chose an immediate reward, then they received the liquid reward immediately. The number of trials in which participants chose delayed reward in the free-choice trials was 3.47 ± 1.58 (mean $\pm$ SD), ranging from 1 to 7.

In order to estimate SVs of delayed rewards, the amount of immediate reward was adjusted based on prior choices in the free-choice trials for each delay condition (10, 30, or 60 s) for each participant (Materials and Methods; Jimura et al., 2009, 2011, 2013). Participants showed delay discounting, with a smaller SV for a longer delay ($t_{(33)} = 2.1$, $p < 0.05$, planned linear contrast; Fig. 3A). The delay discounting for real liquid rewards occurring for delays on the order of seconds was consistent with that in previous reports (Jimura et al., 2009, 2011, 2013). The discounting is also consistent with non-human animal studies based on a procedure using forced and free-choice trials without quantitative information (Richards et al., 1997; Green and Myerson, 2004; Freeman et al., 2009; Vanderveldt et al., 2016; Carter and Redish, 2016).

Dynamics of delay period activity during the forced choice trial

We examined temporal dynamics of brain activity during the delay period of the forced choice trial, where a delayed reward

Figure 3. *A*, Behavioral results. Estimated SVs of a delayed reward as a function of the delay duration. The vertical axis indicates the SV of the delayed reward standardized by its reward amount (12 ml), and a value of 1.0 reflects that the delayed reward is not time discounted. Error bars indicate standard errors of the mean across participants. *B*, Brain regions showing activation dynamics of the AU (left) and upcoming unrewarded reward value (right) during the delay period of the forced choice trial. The color bars indicate significance levels. View direction of the 3D surface of the brain was indicated by the parallelogram and arrow on the right bottom. The green arrow head indicates aPFC focus showing the AU effect. Black and white arrow heads indicate frontal and temporal poles, respectively. *C*, The time course of brain activity demonstrates a gradual decrease toward the reward outcome. *D*, BOLD models of the AU (left) in the first (red) and last (green) trials. Time course of the difference between the first and last trials (right). *E*, BOLD signals in the aPFC showing AU effect (left) in the first (red) and last (green) trials. Time course of the difference between the first and last trials (right).

was experienced for the first time ever, and choice preference was being formed. Importantly, before every forced choice trial, participants were unaware of when they would be given the upcoming future delayed reward, because they had no prior experience with the choice option indicated by a picture that was unique for each trial block (Fig. 1*B,C*; see also above, Behavioral procedures). However, it was possible to anticipate when the future reward would become available, based on the experience of past trial blocks using different pictures to indicate rewards. Thus, the current analysis assumed that participants expected a delay duration before each forced choice trial for the delayed reward. The expectation was then updated through the experiences of the forced choice trials across trial blocks.

In order to quantitatively model this idea, the current analysis used a Bayesian inference learning approach (Behrens et al.,

2007; see Materials and Methods). The expectation of reward outcome was modeled stochastically, based on a probability density distribution of reward outcome occurring in the future. The probability distribution was calculated based on a γ distribution as a function of expected delay duration (Fig. 2*A*). The distribution was then updated every time after participants completed a forced choice trial for the delayed reward.

Given theoretical and empirical predictions in prior work (Berns et al., 2006, 2007; Roesch et al., 2006; Rangel et al., 2008; Jimura et al., 2013; Iigaya et al., 2020), two temporal components of value representation signals were modeled while the future reward was anticipated. One component was the AU model, reflecting the utility (or pleasure) of anticipation of a future reward (Loewenstein, 1987), and the other component was the value of upcoming future reward (Green and Myerson, 2004;

Table 1. Brain regions showing significant effects of AU and upcoming unexperienced reward value during delay period of forced choice trial

Effect	Area	x	y	z	t value	BA
AU	Frontal	36	32	32	9.2	9
		40	20	46	7.9	8/6
		40	54	−14	7.7	11
		28	42	−16	7.4	11/47
		−44	52	−2	5.4	10
		−30	60	−16	5.2	11
		−26	48	−18	5.2	11/47
		−44	46	−16	5.0	10
		−32	28	28	5.0	9/44
		−34	60	8	4.5	10/46
	Parietal	−26	56	−2	4.2	10
		−38	26	42	4.2	8/6
		−38	16	60	3.9	6
		−36	14	40	3.4	9/8
		−50	36	26	3.0	46/9
		−36	−26	54	−12.4	4
		−4	−10	52	−10.6	6
		−44	−54	42	8.9	40
		−50	−50	54	7.9	40
		−32	−60	46	6.8	40/7
	Temporal	−48	−38	44	6.0	40
		−50	22	42	5.5	40/7
		−34	−84	38	−11.4	40
		58	−24	−18	6.0	20
		64	−36	−8	5.8	21
		42	−44	−4	5.2	21
		54	−42	−14	4.5	20/37
		−44	−78	30	−11.8	39
		−58	−66	0	−10.9	21/37
		−40	−22	18	−10.5	43
	Occipital	16	−94	4	10.4	17
		4	−72	50	10.4	7
		−20	−98	8	9.3	19
		8	−90	−8	8.8	18
		−36	−72	−28	11.2	Cerebellum
		−10	−76	−30	8.8	Cerebellum
		−20	−8	24	9.4	Caudate
		−16	4	20	8.2	Caudate
		−4	54	28	7.1	10/9
		38	28	−2	5.6	Insula
Upcoming unexpected Reward value	Frontal	46	32	−14	5.4	47
		32	16	−8	5.1	Insula
		48	24	18	4.7	45/9
		6	52	12	4.6	10
		−46	34	−14	5.2	47
		−32	22	6	4.9	Insula
		−38	34	4	3.7	45/46
		−44	18	−8	3.7	47
		−36	−16	56	5.4	6
		−46	−20	48	3.6	6/4
	Parietal	−40	−26	20	4.2	40
		4	−72	48	−5.8	7
		2	−64	62	−4.6	7
		0	−52	68	−3.3	7
		14	−58	40	−3.3	7/31
		−6	−78	52	−3.2	7
		−12	−58	30	−3.1	31
		−30	−54	32	−5.8	7/39
		−34	−64	54	−4.7	7/39
		−36	−62	40	−4.3	7/39
		−30	−54	48	−3.7	7/39/40
		−26	−44	38	−2.9	7/40
		36	−54	46	−5.1	39/7

(Table continues.)

Table 1. Continued

Effect	Area	x	y	z	t value	BA
	Temporal	36	−66	58	−4.1	7/39
		−62	−36	10	5.9	22
		−60	−18	6	4.8	22
		−48	−40	18	4.3	22
		−36	−40	6	4.2	22/21
		−54	−52	10	4.2	39/21
		64	−34	6	6.2	22/21
		52	−12	−8	5.1	22
		66	−14	−6	3.9	22/21
		68	−24	12	3.8	21/40
	Others	56	−28	−4	3.7	22/21
		60	−18	4	3.6	22
		−38	18	−22	4.6	38
		28	−70	−48	−5.9	Cerebellum
		16	−74	−48	−5.1	Cerebellum
		−6	−72	−28	−4.3	Cerebellum
		28	−48	−42	−4.3	Cerebellum
		18	−66	−34	−4.2	Cerebellum
		8	−94	2	−4.2	Cerebellum
		−18	−48	−32	−4.5	Cerebellum
		−30	−54	−36	−4.4	Cerebellum
		−32	−66	−48	−3.9	Cerebellum
		−42	−54	−30	−3.7	Cerebellum
		−42	−66	−32	−3.5	Cerebellum

Coordinates are listed in MNI space. BA indicates the Brodmann area near the coordinates and is approximate.

Rangel et al., 2008). The AU is maximal when the delay period started and is then gradually decreased as the reward outcome approaches (Loewenstein, 1987), and the value of upcoming future reward showed an inverse dynamics.

In order to model these dynamic components, the PDF was first integrated from the start of the delay to the current time. We integrated the PDF because participants' expectation of reward outcome continues to increase as the delay period is experienced (Green and Myerson, 2004; Janssen and Shadlen, 2005; Berns et al., 2007; Jimura et al., 2013). The accumulated expectation reflects the estimation of the probability that the reward is likely to become available, like hazard rates (Janssen and Shadlen, 2005; McGuire and Kable, 2015). Since the CP is reaching to the upper limit (i.e., 1) as time t goes to infinity, we defined the AU dynamics as 1 minus the CP (Fig. 2B, top; see Materials and Methods).

The inverse dynamics (i.e., CP) may reflect the value of upcoming reward that was not experienced ever but would become available eventually (Fig. 2B, bottom), similarly to prior modeling (Janssen and Shadlen, 2005; Berns et al., 2006; Jimura et al., 2013; McGuire and Kable, 2015). Then, the AU and upcoming reward models were calculated for each forced choice trial for each participant, and the models were convolved with a canonical HRF (Fig. 2C). Parameters were estimated based on a standard GLM approach (see Materials and Methods).

A strong AU effect was observed in multiple regions in the PFC, including fronto-polar regions in the aPFC (Fig. 3B, top; Table 1). On the other hand, the dynamic effect of the upcoming unexperienced reward value is observed in distributed brain regions including medial and lateral prefrontal regions and temporal regions (Fig. 3B, bottom; Table 1).

The time courses of aPFC signals during delay demonstrated the temporal dynamics of AU (Fig. 3C). The AU effect in the aPFC during forced choice trials is consistent with that during current free-choice trials and in prior work (Jimura et al., 2013).

Figure 4. Correlation between AU dynamics and self-control. **A**, Correlation map between AU effect and AuC. Positive correlation indicates greater AU effect in stronger self-controlled individuals (high AuC). White closed line indicates aPFC region showing AU effect during the delay period of the free-choice trial. **B**, A scatter diagram of AuC value and AU effect in aPFC (for display purpose only). Each plot denotes one participant. **C**, **D**, Time courses of brain activity during delay period of forced choice trials. Time courses are shown in separate panels for each group of the AuC value, high (**C**) and low (**D**). **E**, Functional connectivity during the delay period of the forced choice trial. VS activity is predicted by the interaction of aPFC activity and the AuC. Yellow areas indicate ROIs in aPFC defined based on the free-choice trial and in anatomically defined VS.

To evaluate Bayesian learning model reflecting update of reward expectation across trial blocks, we compared temporal dynamics of BOLD models for AU and aPFC signals during the first and last trials (see Materials and Methods). Compared with the last trial, the AU model in the first trial shows a more gradual decrease after the peak (Fig. 3D, left), which is consistent with the economic model during initial 10-s epoch (Fig. 2B, top). The aPFC signals showed similar temporal characteristics, more gradual decrease in the first relative to the last trials after the peak (Fig. 3E, left). The disparity of the models (Fig. 3D, right) and signals (Fig. 3E, right) showed similar temporal characteristics ($r = 0.76$, $p < 0.001$), demonstrating that the current BOLD

model based on Bayesian learning successfully tracked the temporal dynamics of aPFC activity modulated by the experiences of the delayed reward across trial blocks.

Anticipatory aPFC dynamics and self-controlled choice

Prior studies of intertemporal choice have suggested that strong self-control is reflected in less delay discounting (Madden and Bickel, 2009) and the utility of anticipation of future reward (Loewenstein, 1987). We next examined relationships between delay discounting and the AU effect in the aPFC during forced choice trials for the delayed reward, in which choice preference was formed.

Figure 5. Brain activity during consumption in the forced choice trial. **A**, Maps were overlaid on a 2-D slice of anatomic images. Left, Delay effect (delayed reward vs immediate reward). Middle, Free choice effect (chosen vs unchosen). Right, Interaction of the two effects. Black solid closed lines indicate the anatomically defined VS region. **B**, Activity magnitude in anatomically defined bilateral VS regions in each condition. Error bars indicate standard errors of the mean across participants; **** $p < 0.001$, *** $p < 0.005$, ** $p < 0.01$, * $p < 0.05$, Bonferroni-corrected. R: right.

For each participant, the degree of delay discounting was indexed by the AuC of an individual's discounting pattern reflecting the degree of self-control, with greater AuC indicating stronger self-control (Myerson et al., 2001; Sellitto et al., 2010; Jimura et al., 2011, 2013; see Materials and Methods). In order to explore aPFC regions where AU effect was modulated by self-control, voxel-wise intersubject correlations were calculated between parameter estimates of AU dynamics and the AuC measure. We observed significant positive correlation within the aPFC [peak coordinate: (32, 50, -8), 20 voxels, $t_{(33)} = 3.4$, $r = 0.61$; Fig. 4A]. A scatter diagram for this aPFC cluster shows a greater AU effect in individuals with greater AuC (Fig. 4B).

In order to characterize this correlation more specifically, time courses of MRI signals in this aPFC region were extracted for the participants with the highest and lowest tertiles of AuC values (high and low AuC; Fig. 4B). In the high-AuC group, the SVs for 10-, 30-, and 60-s trials were 0.64 ± 0.25 , 0.63 ± 0.14 , and 0.54 ± 0.25 (mean $\pm$ SD; standardized by its amount: 12 ml), respectively. In the low-AuC group, the values were 0.31 ± 0.24 , 0.16 ± 0.11 , and 0.20 ± 0.18 , respectively. In the high-AuC group, the AU effect was evident in all delay period conditions (Fig. 4C), whereas the low-AuC group showed attenuated dynamics (Fig. 4D). These results suggest that more self-controlled choice preferences were formed in individuals showing a stronger AU effect in the aPFC during forced choice trial of delayed reward where participants were anticipating an unexperienced future reward.

Delay-period functional connectivity between the aPFC and VS

Prior studies of value-based decision-making have suggested major roles for the VS in behavioral impulsivity (Tanaka et al., 2004; McClure et al., 2004, 2007; Hariri et al., 2006; Jimura et al., 2013), and the current study demonstrated that the aPFC dynamics during the anticipation of future rewards is associated with the formation of self-controlled choice preferences. We then hypothesized that aPFC dynamics modulated VS activity during the delay period of the forced choice trials in association with the degree of delay discounting. In order to test this hypothesis, we performed a trial-based functional connectivity analysis implemented with a multilevel mixed effect GLM, where the VS

Table 2. VS regions showing significant delay effect, subsequent choice effect, and their interaction effect

Effect	x	y	z	t value
Delay	-14	20	-6	4.0
Choice	-12	20	-2	3.6
Delay $\times$ choice	-12	20	-4	3.5
	10	14	-2	3.2

Coordinates are listed in MNI space.

signal was explained by the aPFC signal during the delay period on a trial-by-trial basis, which was further modulated by individuals' AuC values (see Materials and Methods).

In forced choice trials, the aPFC signal and AuC value was negatively associated (interaction effect: $t_{(16.8)} = -3.2$; $p < 0.01$), suggesting that the VS signal was attenuated by a greater aPFC signal, which was specifically enhanced in strongly self-controlled individuals (aPFC signal > baseline: $t_{(305)} = 8.6$; $p < 0.001$; Fig. 4E).

Reward consumption in the forced choice trial: preference formation effect in VS

Participants directly consumed liquid rewards after the delay period in a delayed reward trial or immediately after the button press in an immediate reward trial. Because the experience of delayed and immediate rewards in the forced choice trials was completed by reward consumption, it is possible that subsequent choices in the free-choice trials were predicted by brain mechanisms involved in reward consumption during the forced choice trials. On the other hand, as shown above, VS activation was reduced by the greater aPFC signal in self-controlled individuals during the delay period of the forced choice trials. Additionally, a number of neurophysiological and neuroimaging studies reported the involvement of VS during intertemporal choice (Tanaka et al., 2004; Kable and Glimcher, 2007; Ballard and Knutson, 2009; Cai et al., 2011; Jimura et al., 2013), reward attainment (Schultz et al., 1992; O'Doherty, 2004), and behavioral impulsivity (Tanaka et al., 2004; McClure et al., 2004, 2007; Hariri et al., 2006; Jimura et al., 2013). This collective evidence suggests the VS region is involved in formation of choice preference during reward consumption.

In order to test this hypothesis, the forced choice trials were classified as a delay (i.e., delayed and immediate rewards) and

Figure 6. *A*, Dynamic models of AU and SV during delay period of the free-choice trial. *B*, *C*, Brain region showing AU and SV effects. *B*, AU effect. *C*, SV effect. Formats are similar to those in Figure 3*B*.

subsequent choice in the free-choice trials (i.e., chosen and unchosen rewards), entailing a 2×2 factorial model, and then activated regions were explored in VS. A main effect of delay was observed in the left anterior VS, with greater activity for delayed reward (Fig. 5*A*, left; Table 2), and a main effect of subsequent choice was observed in the same region, with greater activity for chosen reward (Fig. 5*A*, middle). More interestingly, the VS region also showed an interaction effect of delay and subsequent choice (Fig. 5*A*, right).

To examine the interaction effect more specifically, brain activity during the forced choice trial was extracted from an anatomically defined VS region for (1) chosen delayed reward, (2) chosen immediate reward, (3) unchosen delayed reward, and (4) unchosen immediate reward trials, and compared between trials (Fig. 5*B*; see Materials and Methods). Prominent activity was observed during consumption of chosen immediate reward and unchosen delayed reward (chosen immediate: $t_{(33)} = 3.2$, $p < 0.005$; unchosen delayed: $t_{(33)} = 4.3$, $p < 0.001$). The choice effect (chosen vs unchosen) and reward delay effect (delayed vs immediate) also showed a significant interaction effect ($t_{(33)} = 3.0$, $p < 0.005$). The source of the interaction effect was attributable to reduced VS activity during consumption of chosen delayed reward and unchosen immediate reward, where a self-controlled choice was made (i.e., delayed larger amount of reward was chosen) in the free-choice trial (chosen delayed vs unchosen delayed: $t_{(33)} = -2.8$; $p < 0.01$; unchosen immediate vs chosen immediate: $t_{(33)} = -2.7$; $p < 0.05$). These results suggested that self-controlled choice preference was formed with decreased VS activity during reward consumption.

Dynamics of delay period activity during the free-choice trial

We also examined delay period dynamics in the free-choice trial while participants were waiting for a future reward after they made a choice for a delayed reward. Similarly to the forced choice trials, the two dynamic value components were explored in a whole-brain analysis based on a standard GLM approach. A region in the aPFC showed a strong AU effect (Fig. 6*A*; Table 3) and several occipital areas showing a SV effect (Fig. 6*B*; Table 3), which is consistent with the prior study (Jimura et al., 2013). Notably, the analyses for the free and forced choice trials identified aPFC regions located closely (Fig. 3*B*; Table 1).

Discussion

The current study applied an intertemporal choice paradigm for non-human animals to a human experiment, allowing the examination of neural mechanisms associated with the formation of self-controlled choice preference. In the current intertemporal choice task, the association of reward and picture was unfamiliar at the beginning of each trial block, although approximation of the rewards was possible through the experiences of task blocks. While participants were experiencing a delayed reward for the first time, the aPFC had temporal dynamics reflecting utility of future anticipation, and the aPFC dynamics were enhanced in self-controlled individuals, further reducing VS activity. On the other hand, during consumption of a liquid reward in the forced choice trial, attenuated VS activity predicted self-control choice in the subsequent free-choice trial. These results suggest that aPFC-VS mechanisms during rewarding experience may underlie the formation of self-controlled choice preference.

Dynamic value models while anticipating unexperienced future rewards

In the forced choice trials, while participants were experiencing the delay period, they did not know exactly when the delayed reward would become available, although they were able to expect it to become available in the near future. The current study modeled this expectation based on Bayesian inference learning, assuming that participants' expectation can be explained by a PDF, and the expectation was updated through the experiences of delayed rewards over the entire experiment. The basic idea of this modeling is compatible prior modeling (Behrens et al., 2007), which may also be applicable to animal studies.

In our previous study, while fMRI studies were administered, human participants directly experienced all events occurring in intertemporal choice for real liquid rewards delayed by seconds (Jimura et al., 2013), like the current study. In this previous study, however, quantitative verbal information was used to present reward options, and participants did not experience or learn about the reward options to make a choice, unlike non-human animal studies and the current study. Nonetheless, the AU effect in the aPFC during forced choice trials in the current

Table 3. Brain regions showing significant effects of AU and SV during delay period of free-choice trial

Effect	Area	x	y	z	t value	BA
AU	Frontal	16	30	56	7.3	8/6
		38	24	50	7.3	8/6
		30	54	−14	6.5	47/11
		16	50	36	6.4	9
		24	56	26	6.2	46/9
		−36	46	−4	5.7	10/47
		42	52	0	5.3	10/47
		−34	58	−10	5.0	10/11
		−38	50	16	4.6	10/46
		−24	44	0	4.4	10
		−26	44	−18	3.9	11
		−36	58	6	3.5	10
		−4	−6	46	−7.3	24
		−42	−6	54	−7.0	4/6
	Parietal	−66	−26	12	−5.2	6
		48	−52	46	6.6	40
		56	−40	54	4.1	40
		58	−42	24	3.5	40
		46	−44	60	2.9	7
		2	−30	28	5.9	23
		6	−74	38	5.7	19/7
		12	−56	34	4.8	31/7
		6	−74	54	3.6	7
		−42	−50	36	6.2	40
	Temporal	−54	−50	40	5.7	40
		−46	−56	56	5.1	40/7
		−48	−42	56	3.1	40
		44	−34	20	−7.2	40/42
		−34	−38	52	−6.3	40/7
		58	−24	−20	6.9	20
		66	−36	−8	5.0	21
		54	−38	−10	4.3	21/20
		56	−12	−24	2.9	21
		38	−58	−26	5.6	20/36
	Occipital	−52	−74	20	−6.2	39/22
		−56	−64	−2	−6.1	19/37
		−34	−26	20	−5.9	22/40
		−52	−74	6	−7.6	37/19
		−60	−22	0	−5.1	21/22
		10	−98	−4	5.4	18
	Others	−22	−72	−28	6.3	Cerebellum
		−46	−68	−30	6.1	Cerebellum
		−14	−84	−24	5.9	Cerebellum
		−40	−60	−48	5.9	Cerebellum
		−34	−62	−26	5.4	Cerebellum
		38	−72	−24	4.7	Cerebellum
		10	−82	−24	4.2	Cerebellum
		50	−54	−32	4.2	Cerebellum
		32	−84	−24	3.8	Cerebellum
		50	−68	−26	3.4	Cerebellum
		−24	−16	−16	−6.9	Hippocampus
		−2	12	−2	−6.3	Striatum
		26	−42	−8	3.7	36/37
		−38	−62	−18	3.2	37/19
SV	Temporal	24	−90	18	6.1	19
		26	−72	32	4.2	19/39
	Occipital	34	−94	10	4.1	18/19
		14	−104	10	3.6	18
		−28	−96	12	4.7	18/19
		−26	−84	10	4.1	18/19
		−10	−90	30	3.7	19
		−16	−90	14	3.6	18/19
		−38	−76	−2	3.2	19

(Table continues.)

Table 3. Continued

Effect	Area	x	y	z	t value	BA
		24	−76	−12	4.5	19/37
		30	−64	−10	4.5	19
		16	−86	−6	3.4	18/19
		−28	−66	−8	4.7	19/37
		−26	−78	−12	3.8	19
	Others	−26	−66	−20	4.0	Cerebellum

Formats are similar to those in Table 1.

study is consistent with that during current free-choice trials and in the prior work (Jimura et al., 2013), yet extends those findings in that the aPFC shows AU dynamics in a situation where the participants were experiencing the delayed reward for the first time and unfamiliar about when a future reward would become available.

In order to model the temporal dynamics of UUFV value during the delay period of forced choice trials, we used CP until the present time, whereas prior modeling used delay discounting curve (Jimura et al., 2013). In the prior study, participants were explicitly told when reward became available with quantitative information, which was not the case in the current study. Thus, the prior modeling was not applicable to the forced choice trials of the current study.

The dynamics of brain activity during the delay period may reflect retrospective signals of the visually presented picture indicating the delayed reward, which may gradually decrease through the delay period. Additionally, because participants were instructed to memorize the picture-reward association, it is possible that they actively maintained the picture during the delay period, which may be associated with sustained activity dynamics. The current experimental design does not allow us to dissociate those dynamics and AU dynamics methodologically. However, the visual information of the choice stimulus itself, which reflected the retrospective signal, is unlikely to be associated with the degree of self-control dominantly, because the learning of stimulus-reward association was not completed during the delay period of the forced choice trial. Additionally, the anticipatory dynamics shows a gradual decrease in signal magnitude, which is incompatible with the sustained dynamics derived from the active maintenance of the visual stimulus.

On the other hand, in the free-choice trial, when subject made a choice for a delayed reward, they knew when they received the reward from the prior experience of the forced choice trial although quantitative information was unavailable. To model the temporal dynamics of value components during the delay period of the free-choice trial, we used a linear model rather than a hyperbolic function that was fitted to SV as we did in our prior study (Jimura et al., 2013). In the prior study, this modeling was possible because SV was estimated in a separate behavioral session before a scanning session. In contrast, in the current study, SV was estimated by adjusting the amount of immediate reward throughout the scanning session. Thus, to avoid circular analysis (Kriegeskorte et al., 2009), we used the linear modeling similarly to a prior study (Berns et al., 2006).

aPFC as a region critical for future anticipation and self-controlled preference formation

The aPFC showed temporal dynamics of AU effects during the delay period of free-choice trials, as observed in a prior study (Jimura et al., 2013). However, these two consistent observations

are considered a postchoice effect. The aPFC also showed delay-period dynamics of AU in the forced choice trials before the free-choice trial, which was correlated with the degree of self-control. Although it is still unclear whether this self-control strength is formed during experimental session or before experiment as an intrinsic trait of self-control, the current study demonstrated that the stronger AU effect predicts choice preference toward a larger amount of delayed reward.

VS mechanisms involved in intertemporal choice

The greater aPFC signal specifically observed in self-controlled individuals attenuated the VS signal during the delay period of the forced choice trial, which is consistent with previous findings showing negative functional connectivity from this same aPFC region to VS, with the connectivity predicting weaker trait impulsivity (Diekhof and Gruber, 2010; Jimura et al., 2013). Additionally, greater VS activity during consumption of the reward in the forced choice trials predicted impulsive choice in the free-choice trial, which is also consistent with previous studies suggesting that the VS is associated with behavioral impulsivity (McClure et al., 2004; Hariri et al., 2006). The current VS involvement further extended those previous findings in that the VS is associated with the formation of impulsive choice, and VS function attenuated by enhanced aPFC activity while experiencing a delayed reward for the first time.

Experimental paradigms for human and non-human animals

Standard non-human animal experiments have used the direct experience of primary rewards, and subjects learned reward options before free-choice trials during experimental session (Richards et al., 1997; Green and Myerson, 2004; Freeman et al., 2009; Carter and Redish, 2016; Vanderveldt et al., 2016). On the other hand, standard human experiments have used choices for secondary rewards presented in hypothetical situations, assuming that valuation of the reward was unchanged throughout the entire experiment, and that participants made choices in the same way as they make in the real world (McClure et al., 2004, 2007; Tanaka et al., 2004; Hariri et al., 2006; Kable and Glimcher, 2007; Shamosh et al., 2008; Hare et al., 2009).

Aiming to bridge the gap between non-human and human experiments, prior studies have demonstrated that humans discount delayed reward for real liquid reward delayed by seconds (Jimura et al., 2009, 2011, 2013). However, these studies still presented quantitative information about rewards, the reward amount and duration of delay, which allowed participants to make choices by evaluating options based on the presented quantitative information. In contrast, such decision-making was impossible in the current study, as choice options were presented without quantitative information. Thus, the current experiment is more compatible with non-human animal experiment than prior studies (Jimura et al., 2009, 2011, 2013).

Another compatibility between the current experiment and prior non-human animal experiments is the behavioral procedure to learn choice options. In standard animal experiments, the subjects directly experienced delayed and immediate rewards during forced choice trials, immediately followed by free-choice trials (Green and Myerson, 2004; Freeman et al., 2009; Carter and Redish, 2016; Vanderveldt et al., 2016). The current study applied this procedure, which aligned the human behavioral procedure with the animal experiments, and demonstrated that humans discount delayed rewards to a greater degree with a longer delay.

In prior human studies of liquid reward (Jimura et al., 2009, 2011, 2013), the discounting curve looks less steep than that in the current study, especially for short delay. Such steep discounting in short delays has also been observed in animal studies (cf. Green and Myerson, 2004; Vanderveldt et al., 2016). In both of the human and non-human animal experiments, subjects directly experienced delay without quantitative information, and choice was made based only on preceding experiences. Thus, one possibility for the greater discounting that occurs in short delay is the discrepancy between subjectively experienced and quantitatively exact duration; this discrepancy became greater or smaller depending on the actual duration. The steep discounting in short delay may be attributable to the absence of quantitative information.

Evolution, aPFC, and future anticipation

The current aPFC region is located in the lateral part of Brodmann area 10, a polar region in the frontal cortex. Compared with other prefrontal regions, this region is disproportionately developed in humans (Semendeferi et al., 2001; Petrides et al., 2012). Its thickened supragranular layer enables the area to connect to other higher-level association cortices (Semendeferi et al., 2001), pointing to a possibility that this area produces high-level mental functions that are most evolved in humans. Indeed, this area is associated with integrative reasoning (Christoff et al., 2001; Bunge et al., 2005) and high-level goal representation (Koechlin et al., 2003).

The aPFC is also involved in imagination and/or simulation of events that may occur in the future (episodic future thinking: Atance and O'Neill, 2001; Schacter et al., 2007; Peters and Büchel, 2010; Benoit et al., 2011; for further discussion, see also Jimura et al., 2013). It is presumed that acquisition of the ability of future prospection is one of the remarkable evolutionary changes (Gilbert and Wilson, 2007), and self-controlled behavior involves such future prospection, enabling decision-makers to maximize reward attainment in the future (Madden and Bickel, 2009; Peters and Büchel, 2011). Not surprisingly, the aPFC shows greater engagement in computing prospective value (Doll et al., 2015), restricting current options (Crockett et al., 2013), and choosing an alternative option (Boorman et al., 2011), which are decision strategies that maximize future reward attainments.

Taken together, our findings suggest that the human aPFC plays a critical role in the formation of self-controlled choice preference while humans anticipate a future reward; this was revealed by the application of a standard behavioral paradigm for non-human animals to a human experiment.

References

- Ainslie G (2005) Précis of breakdown of will. *Behav Brain Sci* 28:635–650.
- Atance CM, O'Neill DK (2001) Episodic future thinking. *Trends Cogn Sci* 5:533–539.
- Ballard K, Knutson B (2009) Dissociable neural representation of future reward magnitude and delay during temporal discounting. *Neuroimage* 45:143–150.
- Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. *Nat Neurosci* 10:1214–1221.
- Benoit RG, Gilbert SJ, Burgess PW (2011) A neural mechanism mediating the impact of episodic prospection on farsighted decisions. *J Neurosci* 31:6771–6779.
- Berns GS, Chappelow J, Cekic M, Zink CF, Pagnoni G, Martin-Skurski ME (2006) Neurobiological substrates of dread. *Science* 312:754–758.

- Berns GS, Laibson D, Loewenstein G (2007) Intertemporal choice – toward an integrative framework. *Trends Cogn Sci* 11:482–488.
- Boorman ED, Behrens TE, Rushworth MF (2011) Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLoS Biol* 9:e1001093.
- Bunge SA, Wendelken C, Badre D, Wagner AD (2005) Analogical reasoning and prefrontal cortex: evidence for separable retrieval and integration mechanisms. *Cereb Cortex* 15:239–249.
- Cai X, Kim S, Lee D (2011) Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. *Neuron* 69:170–182.
- Carter EC, Redish AD (2016) Rats value time differently on equivalent foraging and delay-discounting tasks. *J Exp Psychol Gen* 145:1093–1101.
- Christoff K, Prabhakaran V, Dorfman J, Zhao Z, Kroger JK, Holyoak KJ, Gabrieli JDE (2001) Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage* 14:1136–1149.
- Ciric R, Wolf DH, Power JD, Roalf DR, Baum GL, Ruparel K, Shinohara RT, Elliott MA, Eickhoff SB, Davatzikos C, Gur RC, Gur RE, Bassett DS, Satterthwaite TD (2017) Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage* 154:174–187.
- Crockett MJ, Braams BR, Clark L, Tobler PN, Robbins TW, Kalenscher T (2013) Restricting temptations: neural mechanisms of precommitment. *Neuron* 79:391–401.
- Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69:1204–1215.
- Diekhof EK, Gruber O (2010) When desire collides with reason: functional interactions between anteroventral prefrontal cortex and nucleus accumbens underlie the human ability to resist impulsive desires. *J Neurosci* 30:1488–1493.
- Doll BB, Duncan KD, Simon DA, Shohamy D, Daw DN (2015) Model-based choices involve prospective neural activity. *Nat Neurosci* 18:767–772.
- Eklund A, Nichols TA, Knutsson H (2016) Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc Natl Acad Sci USA* 113:7900–7905.
- Figner B, Knoch D, Johnson EJ, Krosch AR, Lisanby SH, Fehr E, Weber EU (2010) Lateral prefrontal cortex and self-control in intertemporal choice. *Nat Neurosci* 13:538–539.
- Frederic S, Loewenstein G, O'Donoghue T (2002) Time discounting and time preference: a critical review. *J Econ Lit* 40:351–401.
- Freeman KB, Green L, Myerson J, Woolverton WL (2009) Delay discounting of saccharin in rhesus monkeys. *Behav Processes* 82:214–218.
- Gilbert DT, Wilson TD (2007) Propection: experiencing the future. *Science* 317:1351–1354.
- Glimcher PW (2009) Choice: towards a standard back-pocket model. In: *Neuroeconomics: decision and the brain* (Glimcher PW, ed), pp 503–522. San Diego: Academic Press.
- Green L, Myerson JA (2004) Discounting framework for choice with delayed and probabilistic rewards. *Psychol Bull* 130:769–792.
- Hare TA, Camerer CF, Rangel A (2009) Self-control in decision-making involves modulation of the vmPFC valuation system. *Science* 324:646–648.
- Hariri AR, Brown SM, Williamson DE, Flory JD, de Wit H, Manuck SB (2006) Preference for immediate over delayed reward is associated with magnitude of ventral striatal activity. *J Neurosci* 26:13213–13217.
- Iigaya K, Hauser TU, Kurth-Nelson Z, O'Doherty JP, Dayan P, Dolan RJ (2020) The value of what's to come: neural mechanisms coupling prediction error and reward anticipation. *Sci Adv* 6:eaba3828.
- Janssen P, Shadlen MN (2005) A representation of the hazard rate of elapsed time in macaque area LIP. *Nat Neurosci* 8:234–241.
- Jimura K, Myerson J, Hilgard J, Braver TS, Green L (2009) Are people really more patient than other animals? Evidence from human discounting of real liquid rewards. *Psychon Bull Rev* 16:1071–1075.
- Jimura K, Myerson J, Hilgard J, Keighley J, Braver TS, Green L (2011) Domain independence and stability in young and older adults' discounting of delayed rewards. *Behav Processes* 87:253–259.
- Jimura K, Chushak MS, Braver TS (2013) Impulsivity and self-control during intertemporal decision making linked to the neural dynamics of reward value representation. *J Neurosci* 33:344–357.
- Kable JW, Glimcher PW (2007) The neural correlates of subjective value during intertemporal choice. *Nat Neurosci* 10:1625–1633.
- Kalenscher T, Pennartz CM (2008) Is a bird in the hand worth two in the future? The neuroeconomics of intertemporal decision-making. *Prog Neurobiol* 84:284–315.
- Keeny RL, Raiffa H (1993) Decisions with multiple objectives: preferences and value trade-offs. New York: Wiley.
- Keerativittayayut R, Aoki R, Sarabi MT, Jimura K, Nakahara K (2018) Large-scale network integration in the human brain tracks temporal fluctuations in memory encoding performance. *Elife* 7:e32696.
- Kirby KN, Petry NM, Bickel WK (1999) Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *J Exp Psychol Gen* 128:78–87.
- Koechlin E, Hyafil A (2007) Anterior prefrontal function and the limits of human decision-making. *Science* 318:594–598.
- Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. *Science* 302:1181–1185.
- Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12:535–540.
- Loewenstein G (1987) Anticipation and the valuation of delayed consumption. *Econ J* 97:666–684.
- Luo S, Ainslie G, Giragosian L, Monterosso JR (2009) Behavioral and neural evidence of incentive bias for immediate rewards relative to preference-matched delayed rewards. *J Neurosci* 29:14820–14827.
- Madden GJ, Bickel WK (2009) Impulsivity: the behavioral and neurological science of discounting. Washington, DC: American Psychological Association.
- McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004) Separate neural systems value immediate and delayed monetary rewards. *Science* 306:503–507.
- McClure SM, Ericson KM, Laibson DI, Loewenstein G, Cohen JD (2007) Time discounting for primary rewards. *J Neurosci* 27:5796–5804.
- McGuire JT, Kable JW (2015) Medial prefrontal cortical activity reflects dynamic re-evaluation during voluntary persistence. *Nat Neurosci* 18:760–766.
- Mischel W, Shoda Y, Rodriguez MI (1989) Delay of gratification in children. *Science* 244:933–938.
- Montague BR, Berns GS (2002) Neural economics and the biological substrates of valuation. *Neuron* 36:265–284.
- Myerson J, Green L, Warusawitharana M (2001) Area under the curve as a measure of discounting. *J Exp Anal Behav* 76:235–243.
- O'Doherty JP (2004) Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr Opin Neurobiol* 14:769–776.
- Ottens LJ, Henson RNA, Rugg MD (2002) State-related and item-related neural correlates of successful memory encoding. *Nat Neurosci* 5:1339–1344.
- Peters J, Büchel C (2010) Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-midtemporal interactions. *Neuron* 66:138–148.
- Peters J, Büchel C (2011) The neural mechanisms of inter-temporal decision-making: understanding variability. *Trends Cogn Sci* 15:227–239.
- Petrides M, Tomaiuolo F, Yeterian EH, Pandya DN (2012) The prefrontal cortex: comparative architectonic organization in the human and the macaque monkey brains. *Cortex* 48:46–57.
- Rachlin H (2004) The science of self-control. Cambridge: Harvard University Press.
- Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci* 9:545–556.
- Raudenbush SW, Bryk AS (2002) Hierarchical linear models, Ed 2. Thousand Oaks: Sage.
- Richards JB, Mitchell SH, de Wit H, Seiden LS (1997) Determination of discount functions in rats with an adjusting-amount procedure. *J Exp Anal Behav* 67:353–366.
- Roesch MR, Taylor AR, Schoenbaum G (2006) Encoding of time-discounted rewards in orbitofrontal cortex is independent of value representation. *Neuron* 51:509–520.
- Rolls ET, McCabe C, Redoute J (2008) Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task. *Cereb Cortex* 18:652–663.

- Schacter DL, Addis DR, Buckner RL (2007) Remembering the past to imagining the future: the prospective brain. *Nat Rev Neurosci* 8:657–661.
- Schultz W, Apicella P, Scarnati E, Ljungberg T (1992) Neuronal activity in monkey ventral striatum related to the expectation of reward. *J Neurosci* 12:4595–4610.
- Sellitto M, Ciaramelli E, di Pellegrino G (2010) Myopic discounting of future rewards after medial orbitofrontal damage in humans. *J Neurosci* 30:16429–16436.
- Semendeferi K, Armstrong E, Schleicher A, Zilles K, van Hoesen GW (2001) Prefrontal cortex in humans and apes: a comparative study of area 10. *Am J Phys Anthropol* 114:224–241.
- Shamosh NA, Deyoung CG, Green AE, Reis DL, Johnson MR, Conway AR, Engle RW, Braver TS, Gray JR (2008) Individual difference in delay discounting: relation to intelligence, working memory, and anterior prefrontal cortex. *Psychol Sci* 19:904–911.
- Stott JJ, Redish AD (2014) A functional difference in information processing between orbitofrontal cortex and ventral striatum during decision-making behaviour. *Phil Trans R Soc B* 369:20130472.
- Tanaka SC, 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.
- Vanderveldt A, Oliveira L, Green L (2016) Delay discounting: pigeon, rat, human—does it matter? *J Exp Psychol Anim Learn Cogn* 42:141–162.