Supplemental Material

Subjects

Twenty-two healthy, right-handed male volunteers, with no history of psychiatric or neurological disorders, gave written informed consent after the nature and possible consequences of the study were explained. The study was approved by the ethics and safety committees of ATR and of Hiroshima University. On the screening day, a psychiatrist interviewed each volunteer and screened them for psychiatric problems using the Structured Clinical Interview for DSM-IV; further, each volunteer had a health examination including blood and urine tests, a chest X-ray, and an electrocardiogram. We recruited only male subjects to avoid estrogen level fluctuation during the menstrual cycle in women, which affects central serotonin levels. We excluded participants who had health and/or psychiatric problems.

We present the results for twenty subjects, as two subjects were excluded from the study. The first subject was excluded because no change in plasma free tryptophan measurements between control tryptophan and high tryptophan conditions could be detected. This can be explained by either an error in the procedure, or by digestive problems, as all other subjects exhibited a dramatic increase in plasma free tryptophan measurements in the high tryptophan condition (close to a 40-fold increase compared to pre-ingestion measurements). The second subject was excluded because of a technical problem that prevented us to record the choice data in the tryptophan depletion condition.

Tryptophan manipulation

Each subject participated in four sessions: a first session for screening and task practice followed by three experimental sessions, each under on of three different tryptophan conditions (depletion, loading, and control). The three experimental sessions were scheduled with a minimal interval of one week between each to remove the effects of tryptophan dietary control induced in the preceding session. The experiment was a within-subject, double-blind, placebo-controlled, and counter-balanced design in which a controller, who was not an experimenter, prepared three types of amino acid mixtures,

which order was randomly scheduled for each subject. To maximize the dietary effect, subjects were instructed to consume a low-protein diet that we provided (less than 35 g/day total) for 24 hours before the experiment, and to fast overnight before each experimental day (Delgado et al., 1990; Bjork et al., 1999; Bjork et al., 2000).

On each experimental day, subjects consumed one of three amino acid mixtures and underwent two venipunctures to determine their plasma free tryptophan concentration, which is known to correlate with the serotonin level in the cerebrospinal fluid (Young and Gauthier, 1981; Young et al., 1985; Delgado et al., 1990; Carpenter et al., 1998; Williams et al., 1999; Bjork et al., 2000). The first blood samples were obtained before consumption of the amino acid mixture to confirm the tryptophan baseline, and the second blood samples were taken six hours after consumption to determine the effect of dietary manipulations on the plasma tryptophan level.

Just before the second blood draw, subjects completed the Profile of Mood States (POMS) (McNair et al., 1971) and The Beck Depression Inventory (BDI) (Beck et al., 1961). The POMS is a self-report, five-point questionnaire that assesses present, subjective mood state, which comprises of the following six subscales: tension-anxiety, depression-dejection, anger-hostility, vigor-activity, fatigue, and confusion-bewilderment. The BDI is a standard self-report measurement to assess depressive symptoms. The behavioral task started just after the second blood draw.

Amino acid mixtures

We prepared amino acid mixtures consisting of the following quantities of 15 amino acids partially dissolved in 350 ml of water: L-tryptophan: 0 g (depletion), 10.3 g (loading), 2.3 g (control), 5.5 g L-alanine, 4.9 g L-arginine, 3.2 g glycine, 3.2 g L-histidine, 8.0 g L-isoleucine, 13.5 g L-leucine, 11.0 g L-lysine monohydrochloride, 5.7 g L-phenylalanine, 12.2 g L-proline, 6.9 g L-serine, 6.5 g L-threonine, 6.9 g L-tyrosine, and 8.9 g L-valine. This aqueous suspension was flavored with 10 ml of chocolate syrup. In addition, 2.7 g L-cysteine and 3.0 g L-methionine were administered in a little water with each of the trp- , trp+ and control drinks due to their unpalatability in the beverage.

The mean plasma free tryptophan concentrations at the time of the experiment in the control condition was 2.42 ± 0.98 SD μg/ml. These levels are slightly higher than normal physiological levels, around 1.3~1.5 μg/ml (Coppen et al., 1972, , 1973; Hoshino et al.,

1986), but much lower those in the tryptophan loading condition (61.2 ± 34 SD μg/ml), and much higher than those in the tryptophan depletion condition (concentrations below the technical limit for measurement of tryptophan of the laboratory (< 0.0613 μg/ml) for all subjects).

Task

Two stimuli (one white coded for the small reward, and one yellow coded for the large reward) were presented during a time selected from a uniform distribution ranging from 0.4 to 0.7 sec from the onset of the presentation of the stimuli. Then, a change of color in the fixation cross from white to red signals acted as a "Go" signal; the subject must then decide to pursue either the large or the small reward. The subject then clicked on the mouse button associated with the position of the chosen stimulus (i.e. left button to choose the left stimulus, for instance). 1.5 second after the beginning of the step, two new stimuli were presented, and a new step starts – the stimulus that was chosen shows more filled patches and the stimulus that was not chosen is identical to that of the previous step. A trial ends when either square is completely filled (100 patches are filled). The corresponding monetary reward then appears on the screen for 1.5 sec (corresponding to an inter-trial interval of 1.5 sec). To maintain the subjects' attention, the position of the squares (left or right) is changed randomly at each step (see Figure 1)

At each trial, the delays to the small and large rewards D_S and D_L are theoretically given by:

$$D_S = (100\text{-}N_S) / S_S * ts \text{ and } D_L = (100\text{-}N_L) / S_L * ts$$

where *ts* is the time step (1.5 sec), N_S and N_L are the initial number of white and yellow patches, and S_S and S_L the number of patches added per step (10 ± 2 patches/step). At the onset of each trial, the white and yellow patches were drawn from random uniform distributions: white patches were in the range 85 ± 10 and initial yellow patches in the range 40 ± 35. Thus, the white square always appeared brighter than the yellow square on the first step of each trial, and the average delay needed to get a large reward was 4 times that that to get a small reward (excluding the inter-trial interval). For the average value of S_S and S_L, the range of theoretical delays for the small rewards was 0.75 to 3.75 sec, and for the large rewards 3.75 to 14.25 sec. Because the experimental step is 1.5 sec, however, the actual delays were the delays above rounded to the next 1.5 sec

increment; further every trial also contained an additional step due to the inter-trial interval *ITI* = 1.5 sec.

In the “dynamics environment”, the number of patches removed per step was slowly and linearly increased from 10 to 24 for the large rewards, and decreased from 10 to 6 for the small rewards from the onset of block 6 to the end of block 9; thus, the estimated values needed to be readjusted to maximize the total gain in this dynamic environment. To maintain favorable small reward choices in spite of these faster removal of patches for the large reward stimulus, the initial ranges of white patches varied linearly as a function of the number of steps from 85 ± 10 to 65 ± 24 and that of yellow patches from 40 ± 35 to 60 ± 23, from the onset of block 6 to the end of block 9. The session was divided in 10 blocks, each lasting 210 seconds and separated by 15 seconds rest time from the next. The first four blocks were used as warm-up blocks to obtain stable subject’s responses and convergence of the value functions estimated by the model (see below).

Reinforcement model fit and statistical analysis

At each step, the action value *Q(s(t),a(t))*, represents the expected sum of discounted future rewards the by taking the action *a(t)* (i.e. choosing the stimulus leading to the large or the small reward) at state *s(t)* (i.e for a particular estimated delay) and following the current policy in subsequent steps

$$Q(s(t),a(t)) = E\left[\sum_{k=0}^{\infty}\gamma^{k} r(t+k)\right], \qquad (1)$$

where *r(t), r(t+1), r(t+2)…* are the rewards acquired by following a certain action policy *P(a|s)* starting from the state *s(t)* and choosing action *a(t)*, and γ is the discount factor. Note that, because $1 \geq \gamma > 0$, future rewards a steps > 1 are more highly discounted than rewards on the next steps. We discretized the delays into five equidistant "states"; the action values *Q* were thus represented by a 2*5 matrix. The temporal difference (TD error), expressed as

$$\delta(t) = r(t+1) + \gamma \max_{a}[Q(s(t+1),a)] - Q(s(t),a(t)), \qquad (2)$$

where $\max_a[Q(s(t+1),a)]$ takes the largest value for each of the two action. This equation compares the predicted gains at step t+1 (the first two terms of the right hand side of the equation) to the current value of the state / action pair (the last term). If the TD error is positive, the estimated action value for the current state action pair was too small, and it should be increased. Conversely, if the agent was too optimistic in its estimated value, it should be decresed. This is achived by using the TD error as the teaching signal to update the value function:

$$\Delta Q(a(t)s(t)) = \alpha\delta(t) \tag{3}$$

where α is the learning rate. Actions are selected by computing the probability to take an action by the soft-max function:

$$P(a_i|s(t)) = \frac{e^{\beta Q(s(t),a_i)}}{e^{\beta Q(s(t),a_1)} + e^{\beta Q(s(t),a_2)}}, \tag{4}$$

where i ∈ [1,2,] and β is the inverse temperature, controls the variability of reward choice. If β is low, the choices are highly stochastic – the agent explores the environment, which is useful to discover rewarding actions. If β is high, there is choice are nearly deterministic, which is useful to maximize gain, after the appropriate actions have been discovered during exploration.

All data points (delays and actions at each steps), except for 1) incomplete trials before breaks and 2) steps before reversals of choice were used in the model fit (such reversals were very rare: 0.58 % of all trials on average). We explored systematically the possible values of the three meta-parameters {α, β, γ} in the ranges [0.01 0.99] with steps of 0.02. We used the data of the first four block of trials to allow initial convergence of the action value function *Q(s(t),a(t))*, initialized at zero. For each subject, each serotonin condition, and each block, we systematically searched for the meta-parameters that yielded the smallest cross entropy error (the negative of the logarithm of the likelihood) between the choices made by the model and that made by the subject. This error is given by:

$$E = -\sum_n \{t^n \log(y^n) + (1 - t^n)\log(1 - y^n)\},$$

where *t* = 1 if it the subject chose a small reward, and *y* is the probability of selecting a small reward as given by the model (see equation 4 in main text).

Results are reported as mean ± standard error. In repeated-measures analysis of variance, the dependent variable was the endpoint (small rewards percentage or individual meta-parameters) and, the factors were the serotonin conditions and blocks. In all repeated measure analyses, only blocks 5 to 10 were included, as the first four blocks were considered as “warming blocks” (see Figure S1). For repeated measures tests, Mauchly’s test of sphericity was used. All the data analyzed in this paper were spherical. Our significance level was $p < 0.05$.

References

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961) An inventory for measuring depression. Arch Gen Psychiatry 4:561-571. .

Bjork JM, Dougherty DM, Moeller FG, Swann AC (2000) Differential behavioral effects of plasma tryptophan depletion and loading in aggressive and nonaggressive men. Neuropsychopharmacology 22:357-369.

Bjork JM, Dougherty DM, Moeller FG, Cherek DR, Swann AC (1999) The effects of tryptophan depletion and loading on laboratory aggression in men: time course and a food-restricted control. Psychopharmacology (Berl) 142:24-30.

Carpenter LL, Anderson GM, Pelton GH, Gudin JA, Kirwin PD, Price LH, Heninger GR, McDougle CJ (1998) Tryptophan depletion during continuous CSF sampling in healthy human subjects. Neuropsychopharmacology 19:26-35.

Coppen A, Eccleston EG, Peet M (1972) Total and free tryptophan concentration in the plasma of depressive patients. Lancet 2:1415-1416.

Coppen A, Eccleston EG, Peet M (1973) Total and free tryptophan concentration in the plasma of depressive patients. Lancet 2:60-63.

Delgado PL, Charney DS, Price LH, Aghajanian GK, Landis H, Heninger GR (1990) Serotonin function and the mechanism of antidepressant action. Reversal of antidepressant-induced remission by rapid depletion of plasma tryptophan. Arch Gen Psychiatry 47:411-418.

Hoshino Y, Yamamoto T, Kaneko M, Kumashiro H (1986) Plasma free tryptophan concentration in autistic children. Brain Dev 8:424-427.

McNair D, Lorr M, Droppelman L (1971) Profile of mood states. San Diego, CA: Educational and Industrial Testing Service.

Williams WA, Shoaf SE, Hommer D, Rawlings R, Linnoila M (1999) Effects of acute tryptophan depletion on plasma and cerebrospinal fluid tryptophan and 5-hydroxyindoleacetic acid in normal volunteers. J Neurochem 72:1641-1647.

Young SN, Gauthier S (1981) Effect of tryptophan administration on tryptophan, 5-hydroxyindoleacetic acid and indoleacetic acid in human lumbar and cisternal cerebrospinal fluid. J Neurol Neurosurg Psychiatry 44:323-328.

Young SN, Smith SE, Pihl RO, Ervin FR (1985) Tryptophan depletion causes a rapid lowering of mood in normal males. Psychopharmacology (Berl) 87:173-177.

Figure S1: Overall time course of a session. Dashed line: Average number of black patches removed per step for the large reward. Solid line: average percentage of small rewards per block for the serotonin control condition. The warm up period allows for the subject's behavior to stabilize and the Reinforcement Learning algorithm to converge. Note how subjects adapt to the "dynamic environment" that favors large rewards (as the number of patches / step removed in the large reward stimulus increased, and simultaneously the number of patches / step removed in the small reward stimulus decreased) by decreasing the preference for small rewards.