bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Motivated for near impossibility: How task type and reward modulates intrinsic motivation and the striatal activation for an extremely difficult task.

Abbreviated title: Motivation for near-impossibility

Kou Murayama [a*, **], Michiko Sakaki[a*], Stef Meliss[a*], Yukihito Yomogida[b, c], Kausu Matsumori[c], Ayaka Sugiura[d], Madoka Matsumoto[e], and Kenji Matsumoto[c*]

[a] School of Psychology and Clinical Language Sciences, University of Reading, Reading, Berkshire, RG6 6AH, UK
[b] Department of Mental Disorder Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Tokyo, 187-8551, Japan
[c] Brain Science Institute, Tamagawa University, Machida, Tokyo, 194-8610, Japan
[d] Department of Pediatrics, Children's Hospital of Michigan, Wayne State University, Detroit, MI 48202, USA
[e] National Institute of Mental Health, National Center of Neurology and Psychiatry, Kodaira, Tokyo, 187-8551, Japan

* Joint first authorship
** Correspondence should be addressed to Kou Murayama, School of Psychology and Clinical Language Sciences, University of Reading, Reading, RG6 6AH, UK and Kenji Matsumoto, Tamagawa University Brain Science Institute, 6-1-1, Tamagawa-gakuen, Machida, Tokyo 194-8610, Japan. E-mail: matsumot@lab.tamagawa.ac.jp.

Conflict of interest statement: The authors declare no competing financial interests.

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Economic and decision-making theories suppose that people would disengage from an extremely difficult task, because such a task does not implicate any normative utility values (i.e. success probability is almost zero). However, humans are often motivated for an extremely challenging task with little chance of success, even without any extrinsic incentives. The current study aimed to address how the nature of the task (luck vs. skill) and the presence of extrinsic rewards modulate this challenge-based motivation, and its neural correlates. Participants played a game-like, skill-based task with three different probabilities of success (i.e., high, moderate, and extremely-low chance of success), in one group without performance-based rewards (no-reward group) and the other group with performance-based monetary rewards (reward group). Participants in the third group played a similar task but the reward outcome was determined in a probabilistic manner (gambling group). Participants in the no-reward group showed increased intrinsic motivation, as the chance of success decreased (i.e. the task becomes more difficult) even if the task was almost impossible to achieve. On the other hand, the reward group exhibited the highest intrinsic motivation when the task had a moderate chance of success, and participants in the gambling group showed decreased intrinsic motivation as the chance of success decreased. The ventral striatum and the ventral pallidum also showed similar patterns of activation. These results suggest that, both at the behavioral and neural levels, people are intrinsically motivated to challenge a nearly impossible task, but only when the task requires certain skills and extrinsic rewards are not available.

Key words: effort, challenge-seeking, reward, ventral pallidum, ventral striatum

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

In December 2015, the Government Communications Headquarters in the United Kingdom released a Christmas puzzle. The puzzle was allegedly extremely difficult and time consuming at the time of announcement, and apparently promised no monetary rewards for solving it. Nevertheless, more than 600,000 people tried to solve the puzzle and (unfortunately) no one succeeded. As illustrated in the example, humans seem to have a natural inclination to challenge an extremely improbable outcome, even without extrinsic incentives (e.g., monetary rewards). It is difficult to explain such challenge-oriented motivation (e.g., Loewenstein, 1999), because there is little chance of succeeding in the task, making normative utility value almost zero according to standard economic, decision-making, and reinforcement-learning theories.

The purpose of the current article is to examine the conditions under which people express such motivation for nearly impossible outcomes and its neural correlates. The dopaminergic reward network is activated more strongly when people are presented with cues signalling higher expected reward value (for reviews Delgado, 2007; Diekhof et al., 2012; Bartra et al., 2013). Based on these findings, the reward network is expected *not* to be activated by cues that predict a task with extremely-low chance of success. Past studies on reward and decision-making, however, overlooked two critical factors for understanding the task motivation for nearly impossible outcomes. First, previous studies mainly employed tasks in which outcome was decided in a pure probabilistic manner (luck task). In contrast, we are often motivated for challenging tasks in our daily life when the task requires some skills. When task success is contingent upon one's skills, people focus more on the improvement of their skills, rather than task success itself; as a result, people may exhibit greater persistence after failure (Lee and Kim, 2014; Lee and Reeve, 2017), perhaps even when the task is almost impossible to achieve.

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Second, previous research has typically implemented extrinsic rewards to quantify the value of task cues. However, humans are often intrinsically motivated for a challenging task in the absence of extrinsic incentives (Deci and Ryan, 1985; Csikszentmihalyi, 1990). In fact, performance-contingent extrinsic rewards can put people in a conflicting situation in the face of a challenging task. Specifically, people's intrinsic motivation for challenge may be offset by the risk of losing out potential extrinsic rewards. In other words, there may be a qualitative difference in people's motivation when extrinsic reward is at stake and when it is not (Murayama et al., 2010a; Baranes et al., 2014; Braver et al., 2014).

We examined whether the nature of the task and the presence of extrinsic rewards influence people's motivation for nearly impossible outcomes and how these factors influence neural responses to the task with different chance of success (including a near impossible task). Participants were engaged in a game-like task (Figure 1A) in which rewarding outcome is determined by participants' skills (skill task) or is probabilistically determined (luck task). Some participants performed the skill task without performance-contingent extrinsic rewards (no-reward group), but others played the task with outcome-contingent monetary rewards (reward group). The remaining participants played the luck task, where they simply received extrinsic rewards that were determined in a pure probabilistic manner (gambling group). Across all groups, each trial was preceded by a cue indicating the success probability of the upcoming task --- high-, moderate-, and extremely-low chance (of success) conditions. Previous studies have shown that the reward network in the brain (e.g. the striatum) represents the rewarding value as well as the motivational relevance/mental preparation for an upcoming task (e.g., Carter et al., 2009; Krebs et al., 2011). Thus, we primarily focus on the activation in the reward network, especially in the striatum.

We predicted that participants would exhibit motivation for the task with extremely-low chance of success only when they were engaged in a skill task without extrinsic

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

incentives (no-reward group; Figure 1B, blue line). On the other hand, when participants were engaged with extrinsic rewards (reward group), we expected that participants' motivation for improbable outcomes would be hampered due to the low expected performance reward (Figure 1B; blue and red lines). When participants were engaged in a luck task with extrinsic incentives (gambling group), there was no room for people to focus on the improvement of their skills. As such, we predicted that participants' motivation would monotonically decrease as the chance of winning rewards decreased, simply because a task with higher chance of success involves higher expected performance reward (Figure 1B, red line). We further expected that these differential patterns of motivation between groups would track the striatal activation in the brain.

Materials and Methods

Participants

Fifty-five participants (mean age = 20.05, *SD* = 1.37; 28 males) were allocated to the no-reward group, the reward group, or the gambling group, constituting a 3 (group, between subjects: no-reward, reward, or gambling) x 3 (chance of success, within-subjects: high chance, moderate chance, or extremely-low chance) factorial design. Participants in the reward and no-reward groups were randomly assigned to these groups, and the data for the gambling group were later collected. Prior to data analyses, we excluded data from four participants: one due to a technical problem, another due to an incidental finding, one due to large motions, and the other because the participant failed to press buttons in most trials. This resulted in 17 participants in the no-reward condition, 17 participants in the reward condition, and 17 in the gamble condition. Thus, the final study sample consisted of *N* = 51 adults aged between 18 and 24 (*M* = 20.1, *SD* = 1.39), with 23 males and 28 females.

Experimental Procedures

Participants played a stopwatch task, which was designed based on the task used in

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

previous research (Murayama et al., 2010b; Murayama et al., 2015). Participants' task was to stop a stopwatch within a specific time window --- between 49.95 and 50.05. There were three types of stopwatches which are different in terms of the chance of success: a) in the high-chance condition, the stopwatch started from 49.95 and ran 100-time slower than an ordinary watch; b) in the moderate-chance condition, the stopwatch started from 45.00 and ran in the same speed as an ordinary watch (like the ones used in previous research); c) in the extremely-low chance condition, the stopwatch started from 0.00 and ran 10-time faster than an ordinary watch. Thus, while the initial time was different across the three types of the stopwatches, they were different in their speed, allowing us to manipulate task difficulties while making the trial duration similar across the conditions (e.g., in all conditions participants needed to press a button around 5 seconds after the stopwatch started in order to stop the stopwatch at 50.00).

In the no-reward and the reward groups, each trial (Figure 1) started with three question marks. Upon participants' button press, the question marks were replaced by a word cue indicating the task difficulty level (i.e. chance of success) of the subsequent stopwatch: "easy" (i.e. high chance of success), "intermediate" (i.e., moderate chance of success), and "super advanced" (i.e. extremely-low chance of success). The cues (and the following stopwatches) were differently colored depending on the chance of success (blue, green, and orange), and the assignment of the colors was counterbalanced across participants. Participants were then shown the stopwatch, and asked to press a button with the right thumb to stop it right between 49.95 and 50.05. The stimulus onset asynchrony (SOA) of the cue presentation and start of the stopwatch task was jittered between 4.5 and 8.5 seconds. Immediately after they pressed the button, they were shown success feedback or failure feedback for 0.8 seconds, followed by a jittered ITI between 3 and 7 seconds. This task sequence was optimized to examine the brain activation associated with the cue presentation.

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Unbeknownst to participants, in the extremely-low chance condition, the displayed time when they stopped the stopwatch was manipulated so that participants always failed. In the moderate chance condition, we made an online adaptive adjustment of task difficulty based on participants' previous task performance so that most participants achieved approximately 50% accuracy. In the high-chance condition, the task was so easy that participants succeeded in all trials without any manipulations. No participants indicated the possibility that their task performance had been manipulated after the experiment. If participants pressed a button too early or too late (e.g., if participants did not press a button within 10 secs after the stopwatch started), a warning message "please press a button" appeared and this trial was skipped.

Prior to the session, participants were told about the differences of the three different types of stopwatches. Specifically, participants were told that the stopwatch game in the high chance condition would be very easy and they can easily win; that the moderate chance condition would be moderately difficult; and that the extremely-low chance condition would be extremely difficult and would be almost impossible to win. Participants in the reward group were further told that they would earn 100 yen (approximately USD 1) for every success irrespective of the type of stopwatch. Participants in the no-reward group were not told anything about performance-based monetary rewards.

The procedure in the gambling group was similar to that in the reward group with a few exceptions. As in the reward group, participants were told that they would earn 100 yen for each time that the stopwatch stopped between 49.95 and 50.05. However, unlike the reward group, they did not have control over the timing at which the stopwatch stopped and the stopwatch was called a lottery machine. They were simply asked to watch the stopwatch starting and stopping on its own and press the button after the stopwatch stopped. If the stopwatch stopped within the time window, this trial was regarded as hit (success) and participants earned money. Participants were told that the outcome would be determined by

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the computer and the task cue would indicate the chance of hit (success) in that trial. Specifically, they were instructed that it would be almost certain that they can earn money in the high-chance condition, whereas it would be almost impossible that they would earn money in the extremely-low chance condition. They were also told that there would be a fair chance (50%) of earning money in the moderate chance condition.

In all three groups, participants also completed a control condition, in which they simply viewed a stopwatch starting and stopping on its own and pressed the button after the stopwatch stopped without any monetary rewards (watch-stop condition). The watch-stop condition was preceded by a word cue “passive viewing”, and the color of the cue and the stopwatch was always gray. The stopwatch ran in one of three different speeds (randomly chosen), which were matched to one of the three chance-of-success conditions). If participants did not press a button within 2 seconds after the stopwatch stopped, a warning message “please press a button” appeared and this trial was repeated.

Across all groups, participants completed 80 trials (20 trials with high chance of success, 20 trials with moderate chance of success, 20 trials with extremely-low chance of success, and 20 watch-stop trials) that were divided into two runs of 40 trials. The sequence of the trials (i.e. the trial order of conditions) was randomized across participants, but to control for any potential trial order effects, we used the same set of trial sequences across the three groups. In addition, the sequence of success and failure trials in the gambling group was matched to that in the reward group. After the session, they completed a self-reported questionnaire asking their intrinsic motivation to engage in the three types of stopwatch tasks (three items taken from Elliot and Harackiewicz, 1996; e.g., "It was fun to play the stopwatch of high chance of success", Cronbach's alpha = .82 - .95). Participants also answered two questions about their perceived rewarding value towards the task cue for each of the three stopwatches (“The cue for the moderately difficult stopwatch made me pleased”, “The cue

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

for the extremely difficult stopwatch made me disappointed", reverse coded, *r*s = .67 - 89).

fMRI Acquisition and Preprocessing

The functional imaging was performed on a 3T magnetic resonance imaging (MRI) scanner (MAGNETOM Trio, A Tim System, Siemens, Germany) with a 32-channel matrix head coil at the Tamagawa University with gradient echo T2*-weighted echo-planar imaging (EPI) sequence. The imaging parameters were TR = 2500 ms, TE = 25 ms, slice thickness = 3 mm, and flip angle = 90°. The data were preprocessed using Statistical Parametric Mapping 12 (SPM12, http://www.fil.ion.ucl.ac.uk/spm/software/spm12/). Images were motion-corrected with realignment to the first volume of the session, co-registered to the bias-corrected, segmented structural image, spatially normalised to the standard Montreal Neurological Institute (MNI) EPI, and spatially smoothed using a Gaussian kernel with a full width at half-maximum (FWHM) of 8 mm.

fMRI Data Analysis

For each participant, the blood oxygen level-dependent (BOLD) response was modeled with the general linear model (GLM) for the following regressors of interest: the high-chance stopwatch cue, the moderate-chance stopwatch cue, the extremely-low-chance stopwatch cue, and the three watch-stop cues (i.e. watch-stop cues with three different speeds; they were separately modeled and then averaged). In addition, feedback for stopwatch trials, error trials (i.e. participants did not press a button for a certain duration; see experimental procedures; for these trials, the entire duration of the trial was modeled), session effects, and motion parameters were included as regressors of no interest. The regressors (except motion parameters and session effects) were calculated using a boxcar function for each stimulus convolved with a canonical hemodynamic response function (HRF) without derivatives. Temporal autocorrelation was accounted for by using a first-order autoregressive model during Classical (ReML) parameter estimation. Results for the three watch-stop cues

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were averaged and used in the following three contrasts to examine the effects of different cues: (i) a contrast between the low-chance stopwatch cue and the averaged watch-stop cues; (ii) another contrast between the moderate-chance stopwatch cue and the averaged watch-stop cues; and (iii) a contrast between the extremely-low chance stopwatch cue and the averaged watch-stop cues.

The resulting contrast images were submitted to a 3 (chance of success: high chance, moderate chance, or extremely-low chance) x 3 (group: no-reward, reward, or gambling) mixed ANOVA with chance of success as within-subject factor and group as between-subject factor. Our primary analysis focused on the reward network in the brain, especially the striatum/pallidum. Thus, we performed a region of interest (ROI) analysis. The striatum/pallidum ROI mask included the bilateral caudate, putamen, and pallidum from the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). For the purpose of completeness, we also conducted a set of whole-brain GLM analyses. We applied a voxel-level family-wise error (FWE)-corrected threshold ($P_{FWE} < 0.05$) for ROI analysis. For the whole-brain analysis, we used cluster-extent FWE-corrected threshold with the cluster defining threshold of $p = 0.001$ ($P_{FWE} < 0.05$).

Parametric Modulation Analysis

We conducted a parametric modulation analysis to examine the brain areas associated with participants' self-report ratings (intrinsic motivation for the task and rewarding value toward the task cue) for each level of chance of success. Specifically, we ran GLMs in which task cues (high, moderate, and extremely-low chance of success) were treated as a single regressor and the mean ratings of intrinsic motivation for the task or the rewarding value toward the task cue were added as a parametric modulator. Other regressors were the same with the main analysis. The contrast of interest was activation associated with the parametric

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

modulator of stopwatch cues compared to the watch-stop cue. For thresholding, we used the same FWE correction methods as in the main analysis.

Functional Connectivity Analysis

We also conducted a functional connectivity analysis (i.e. beta series correlation; Rissman, Gazzaley, & D'Esposito, 2004) in order to examine which brain regions were co-activating with the ventral striatum/ventral pallidum during the cue presentation. To perform this analysis, a separate GLM was constructed in which every stage of every trial was modelled with a separate covariate. We then estimated the functional connectivity (i.e. correlation across estimated beta values in the GLM) between a seed region and other brain areas for the stopwatch and the watch-stop task separately for each chance of success for each subject. The seed regions were 5 mm-spheres around the peak voxel observed in the right and the left ventral striatum/ventral pallidum in the previous ROI analysis. The analyses were performed using the BASCO toolbox (Göttlich, Beyer, & Krämer, 2015), and the ROI masks were created using MarsBar (Brett, Anton, Valabregue, & Poline, 2002). The correlation maps for the watch-stop task were subtracted from those for the stopwatch task, and the resultant maps were compared between the conditions. We used a cluster-extent FWE-corrected threshold ($P_{FWE} < 0.05$) with the cluster defining threshold of $p = 0.005$, and further applied a Bonferroni correction for two separate seed regions (i.e. the left and right ventral striatum/ventral pallidum).

Results

Behavioural results

Ratings of the intrinsic motivation for the task were analysed using a 3 (group: no-reward, reward, or gambling) x 3 (chance of success: high chance, moderate chance, or extremely-low chance) mixed ANOVA. This ANOVA showed a main effect of chance of success, $F(2,96) = 10.02$, $P < .001$, generalized $\eta^2 = 0.13$, which was further qualified by the

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

interaction between group and chance of success, $F(2,96) = 22.72$, $P < .001$, generalized $\eta^2 = 0.41$. The pattern of the results confirmed our prediction (Figure 2A). Specifically, participants in the no-reward group showed a linear pattern with an *increase* in intrinsic motivation when the chance of success *decreased*, exhibiting greatest intrinsic motivation for improbable outcomes. In the reward group, on the other hand, intrinsic motivation was highest in the moderate chance condition; participants exhibited an increase in intrinsic motivation in the task with moderate chance of success compared to the task with high chance of success, but then intrinsic motivation decreased as the chance of success became extremely low; this resulted in an *inverted-U* trend as a function of chance of success. Finally, as predicted, participants' intrinsic motivation in the gambling group *decreased* as the chance of success *decreased*, showing an opposite linear pattern from that in the no-reward group. A post-hoc trend analysis supported our observation of a positive linear trend in the no-reward group, a quadratic trend in the reward group and a negative linear trend in the gambling group, $b = 0.877$, $t(38.70) = 6.55$, $P < .001$.

A similar ANOVA on the ratings of rewarding value toward the task cue also showed a main effect of chance of success, $F(2,96) = 16.73$, $p < .001$, generalized $\eta^2 = 0.21$, which was again qualified by an interaction effect, $F(2,96) = 20.80$, $P < .001$, generalized $\eta^2 = 0.39$. Interestingly, the pattern was somewhat different from intrinsic motivation (Figure 2B). Participants in the no-reward and gamble groups followed the same trends as observed in the ratings for intrinsic motivation, while those in the reward group showed a *decreasing* trend as chance of success decreased: when the task became difficult (i.e. chance of success decreased), participants reported reduced reward value for the task cue. In a subsequent a post-hoc trend analysis, this pattern (positive linear in no-reward, negative linear in reward and gambling) was confirmed, $b = 1.314$, $t(51) = 7.24$, $p < .001$ (see Figure 2B). In both of

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the ANOVAs, the main effect of group was not significant, $Fs(2, 48) \leq 2.33$, $Ps \leq .10$, generalised $\eta^2 s \leq 0.02$.

We also examined participants' task performance. Although we manipulated performance feedback (see Experimental Procedures; this ensures that participants' subjective feelings of success were controlled across participants and groups), it was still possible to examine the objective performance regarding whether participants were able to press the button in the right timing, given that participants in the reward and no-reward groups were motivated to stop the stopwatch right at 5.00 seconds after the start of the stopwatch. We performed a 2 (chance of success; moderate chance vs. extremely-low chance) X 2 (group; no-reward vs. reward) ANOVA on the actual timing participants pressed a button. There was a significant main effect of the group, $F(1, 32) = 4.79$, $P < 0.05$, generalized $\eta^2 = 0.06$ suggesting that participants in the reward group stopped the stopwatch relatively earlier and less accurately ($M = 4.919$sec, $SD = 0.065$) than those in the no-reward group ($M = 4.945$sec, $SD = 0.042$ in the no-reward group). These results are consistent with the incentive-dependent performance decrements reported in previous research (Mobbs et al., 2009); pressures arising from performance-based extrinsic rewards may have impaired task performance. The main effect of chance of success and the interaction effect were not statistically significant ($Ps > .05$).

Main GLM analysis of fMRI data

The 3 (group) x 3 (chance of success) mixed ANOVA in our anatomical ROI (i.e., the striatum) revealed a significant interaction between group and chance of success in the bilateral ventral striatum extending into the ventral pallidum (VS/VP; Figure 2C; $Ps < .05$, voxel-level FWE corrected; $k = 22$ and 26 for the left and the right VS/VP, respectively; voxels in the right cluster survived with voxel-level FWE even at the whole brain level; $P < .05$). Confirming our hypothesis, the pattern of the results was consistent with the self-

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reported intrinsic motivation: whereas participants in the gambling group showed *decreased* activation in the VS/VP for task cues predicting lower-chance task of success, participants in the no-reward group showed *increased* activation for task cues predicting lower-chance of success. Participants in the reward group, on the other hand, showed an inverted U pattern, showing the highest activation for the task cues predicting moderate chance of success.

To demonstrate that these results do not simply reflect the general activation pattern in the entire brain, we also conducted a whole brain analysis to explore other brain areas showing a similar pattern. In this analysis, we defined a contrast that is aligned with the interaction pattern observed in the VS/VP: high chance (of success) = -1, moderate chance = 0, and extremely-low chance = 1 for the no-reward group; high chance = -1; moderate chance = 2, and extremely-low chance = -1 for the reward group; and high chance = 1, moderate chance = 0, and extremely low chance = -1 for the gambling group. The results showed that, in addition to the cluster peaked at the bilateral VS/VP, a large cluster in the bilateral occipital cortex and clusters in the cerebellum, the left inferior parietal lobule, and the bilateral middle/inferior frontal gyrus showed significant activation (Table S1; $Ps < 0.05$, cluster-level FWE corrected). However, the VS/VP cluster exhibited the highest peak activation. For completeness, we also report the full results of the whole brain analysis with the full factorial 3 X 3 ANOVA in Table S2.

GLM using individual reports as parametric modulators

One interesting observation from the previous analyses is that, whereas participants in the no-reward group exhibited the highest intrinsic motivation (or the highest VS/VP activation) with extremely-low chance of success, participants in the reward group showed decreased intrinsic motivation (and the striatal activation) for the cue for extremely-low chance of success relative to the cue for moderate chance of success. One possible explanation is that, in the reward group, participants' intrinsic motivation for the near

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

impossible task was compromised by the fact that they are very unlikely to obtain monetary rewards (Figure 1B). In fact, perceived rewarding value for the task cue linearly decreased in the reward group as the chance of success decreased, suggesting that participants expected less monetary rewards for the task with less chance of success (Figure 2B).

Because intrinsic motivation for the task and expected extrinsic rewards are in direct conflict in the reward group, the group allows us to dissociate the neural correlates underlying these two processes. For that purpose, we conducted a parametric modulation analysis for participants in the reward group, using the ratings for intrinsic motivation for the task or rewarding value toward the task cue at each chance of success as a parametric modulator (Table 1). Consistent with the previous GLM results using ANOVA (Figure 2C), the VS/VP activity was positively correlated with the ratings of intrinsic motivations ($P < .05$, voxel-level FWE corrected). In addition, outside of the striatum ROI, the inferior frontal lobule was positively correlated with intrinsic motivation ($P < .05$, cluster-level FWE corrected). In contrast, the inferior parietal gyrus, the supplementary motor areas, and the precentral gyrus were positively correlated with the ratings of rewarding value for the task cue ($Ps < .05$, cluster-level FWE corrected). Figure 3 shows these overlapping as well as distinct activation patterns for intrinsic motivation and rewarding value for the task cue. These results suggest distinct neural mechanisms underlying intrinsic motivation for the task and the expected value for extrinsic rewards.

Functional Connectivity Analysis

To further examine the neural mechanisms underlying the decreased intrinsic motivation under the extremely-low chance of success in the reward group, we conducted a beta-series functional connectivity analysis (Rissman et al., 2004), which directly compared the functional connectivity in the extremely-low chance condition between the no-reward group and the reward group. The results showed that the activation in the left VS/VP (one of

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the seed regions) was less correlated with the activation in the supplementary motor area in the reward group (relative to the no-reward group). These results suggest there was functional *decoupling* between the left VS/VP and the supplementary motor area when participants in the reward group expected a task with little chance of success (Figure 4, $P < .05$ with Bonferroni correction, cluster-level FWE corrected).

Discussion

The current study aimed to examine the factors influencing people's motivation for a nearly impossible task and its neural correlates. Our results showed that, when the task success depended on participants' skills (i.e. skill task) and did not promise any extrinsic rewards, participants' intrinsic motivation increased as the task became more difficult (i.e. lower chance of success), even to the extent that the task was almost impossible to achieve. However, when the task was skill-based but participants were promised performance-based monetary rewards, participants' intrinsic motivation was hampered when the chance of success was extremely low; their motivation was peaked when the chance of success was moderate, exhibiting an inverted-U shape relationship between chance of success and intrinsic motivation. Also, when the task did not involve any skills and the outcome was determined in a purely probabilistic manner (i.e. luck task), the task motivation monotonically decreased as the chance of success decreased. VS/VP activation in response to the task cue showed the same pattern, and indeed it was associated with the self-reported intrinsic motivation on a trial-by-trial basis. Thus, the present research suggests that, both at the behavioral and neural levels, people are often intrinsically motivated to challenge a nearly impossible task but this happens mainly when the task requires certain skills and extrinsic incentives are not available.

Our findings pose some challenge to the normative economic and decision-making theories, because such a nearly impossible task should have little utility value given the

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

extremely low expected success probability (Loewenstein, 1999). In fact, in the current experiment, participants never succeeded in the task of extremely-low chance of success. Using skill-based tasks, some previous neuroimaging studies have examined the effects of task difficulty by manipulating task demands (Croxson et al., 2009; Kurniawan et al., 2010; Krebs et al., 2011; Skvortsova et al., 2014; Ulrich et al., 2014; Dobryakova et al., 2017). In these studies, however, participants in high demand task conditions exhibited decent task performance, making it difficult to disentangle the pure challenge-oriented motivation from the expected positive outcomes. In addition, most of the previous studies introduced extrinsic rewards to motivate participants toward an effortful task (for exceptions, see (Krebs et al., 2009; Boehler et al., 2011)). As our findings suggest, the presence of performance-based extrinsic rewards would make the interpretation of the results difficult, because such rewards provide external incentive to engage in a task with higher chance of success, which could be in direct conflict with people's intrinsic inclination to work on a task of low chance of success.

Some existing theories have attempted to explain people's challenge seeking behavior. For example, in the field of robotics and computational modelling, Oudeyer et al. (2013) (see also Baranes et al., 2014) argued that, when there are no explicit extrinsic rewards, agents may use learning progress as intrinsic rewards to motivate behavior. In the literature of psychology, researchers indicated that people seek (moderate) challenge because this provides people with feelings of competence and control (Nicholls, 1984; Deci and Ryan, 1985; Csikszentmihalyi, 1990), feelings of task achievement and pride (McClelland et al., 1976; Sedikides and Strube, 1997), self-definition (Gendolla and Richter, 2010), and opportunity to reduce uncertainty to gauge one's own ability (Trope, 1979). All of these theories, however, do not provide a viable explanation of why people are motivated for a nearly impossible task, because these theories consider occasional experiences of success as a

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

critical source of motivation. One potential explanation for the observed intrinsic motivation for the nearly impossible task is that participants were motivated to see how close they could be to succeeding in the task. In other words, in the absence of extrinsic rewards, participants might have gauged their achievement and progress against these internally generated sub-goals, rather than the externally-defined task success. In fact, in the current task, participants were able to see how close their button press was to the success. Future research should examine the exact psychological and neural mechanisms underlying the motivation for nearly impossible tasks.

This point also highlights the importance of distinguishing between skill tasks and luck tasks in decision-making and neuroscience research (see also Vostroknutov et al., 2012). A luck task (e.g. number guessing task) is a popular and predominant paradigm in decision science/neuroscience to examine how people or animals learn expected reward value from a task and make a decision. However, as noted above, people and animals also have the capacity to acquire rewarding value from the feeling of competence, acquiring knowledge, and achieving self-set goals (Deci and Ryan, 1985; Gottlieb et al., 2013; Murayama, 2019). This means that more complex decision-making mechanisms might be operative with a task that entails some sort of skill learning or knowledge acquisition (e.g., Knutson et al., 2005; Roiser et al., 2010; Aarts et al., 2014). This aspect of the decision-making process is still underexamined in the literature (for an exception, Baranes et al., 2014), but worth attention for future research.

Parametric modulation analysis in the reward condition showed that the expected monetary reward value toward the task cue and intrinsic motivation for the task were supported by the common as well as distinct brain areas, suggesting the importance of dissociating these subjective states when extrinsic rewards are provided to participants. Among the brain areas uniquely associated with the rewarding value toward the task cue, the

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

inferior parietal lobule is often activated when anticipating an extrinsic reward (Liu et al., 2011), and therefore, one straightforward interpretation is that this activation simply reflects the expected value for extrinsic rewards. Another possibility is that the inferior parietal lobule is related to higher-order comparison and integration of different types of tasks and rewards. In fact, in the reward group, the expected value for monetary rewards was in conflict with the intrinsic value for the difficult task, and participants had to accommodate these opposing values. Supporting this idea, a previous fMRI study showed that the inferior parietal lobule is related to expected large future rewards at the cost of immediate losses (Tanaka et al., 2004), suggesting the possibility that this area is activated when different sources of rewards are in conflict and thus need to be integrated (see also Wisniewski et al., 2015).

Our functional connectivity analysis suggests that there is decreased functional connectivity between the (left) VS/VP and the supplementary motor area when participants in the reward group expected a near impossible task. Supplementary motor area and its adjacent area, presupplementary motor area, have been implicated in internally-generated actions and sense of agency (Gallagher, 2000; Nachev et al., 2008), and has been activated when participants mentalized voluntary engagement (Lee and Reeve, 2013). Importantly, both presupplementary motor areas and the striatum showed *deactivation* when extrinsic incentives undermined people's voluntary engagement in a task (undermining effect; Murayama et al., 2010a). These results indicate that the supplementary motor cortex may play a critical role in initiating voluntary task engagement after receiving the signal of internal rewarding value from the striatum. Although exploratory, our findings may add more evidence to this line of perspective --- the decoupling between the VS/VP and supplementary motor area in the reward condition may represent the disruption of this coordination because there is a direct conflict between extrinsic rewards and intrinsic motivation for challenge.

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The current study demonstrated that people have the inherent capability to motivate themselves for a nearly impossible task without extrinsic rewards, but in reality, failure is a major source of psychological dis-functioning. This is because failure in real life involves a multitude of factors that boost the negative implication of failure, such as social comparison, loss of future opportunities, and stigmatization. Previous studies have identified a number psychological underpinnings that are related to resilience and perseverance to failure (e.g., Duckworth et al., 2007; Trautwein et al., 2009; Yeager et al., 2016) and recent neuroimaging work also started paying attention to the topic (e.g., Bhanji and Delgado, 2014; Murayama et al., 2015; Sinha et al., 2016). Our work should be extended to accommodate such real-life factors, to better understand the potential and limitations of our intrinsic motivation to challenge very difficult tasks.

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgments

This work was supported by MEXT KAKENHI on Innovative Areas to K. Matsumoto (JP24120717) and K. Murayama and M. Sakaki (JP16H06406), JSPS KAKENHI to M. Sakaki (JP16H02053), the Tamagawa University Global COE Program "Origins of the Social Mind" from MEXT, Brain/MINDS from AMED,F. J. McGuigan Early Career Investigator Prize from American Psychological Foundation (to K. Murayama), the Marie Curie Career Integration Grants from European Commission (CIG630680 to K. Murayama & CIG618600 to M. Sakaki) and Leverhulme Trust Research Leadership Award (RL-2016-030, to K. Murayama).

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Aarts E, Wallace DL, Dang LC, Jagust W, Cools R, D'Esposito M (2014) Dopamine and the cognitive downside of a promised bonus. Psychological science 25:1003-1009.

Baranes AF, Oudeyer P-Y, Gottlieb J (2014) The effects of task difficulty, novelty and the size of the search space on intrinsically motivated exploration. Frontiers in Neuroscience 8.

Bartra O, McGuire JT, Kable JW (2013) The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. NeuroImage 76:412-427.

Bhanji Jamil P, Delgado Mauricio R (2014) Perceived Control Influences Neural Responses to Setbacks and Promotes Persistence. Neuron 83:1369-1375.

Boehler CN, Hopf J-M, Krebs RM, Stoppel CM, Schoenfeld MA, Heinze H-J, Noesselt T (2011) Task-Load-Dependent Activation of Dopaminergic Midbrain Areas in the Absence of Reward. The Journal of Neuroscience 31:4955-4961.

Braver TS et al. (2014) Mechanisms of motivation-cognition interaction: Challenges and opportunities. Cognitive Affective & Behavioral Neuroscience.

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox. In *8th International Conference on Functional Mapping of the Human Brain*. Sendai, Japan.

Carter RM, MacInnes J, Huettel S, Adcock RA (2009) Activation in the VTA and nucleus accumbens increases in anticipation of both gains and losses. Front Behav Neurosci 3.

Croxson PL, Walton ME, O'Reilly JX, Behrens TEJ, Rushworth MFS (2009) Effort-Based Cost–Benefit Valuation and the Human Brain. The Journal of Neuroscience 29:4531-4541.

Csikszentmihalyi M (1990) Flow: The Psychology of Optimal Experience. New York, NY: Harper and Row.

Deci EL, Ryan RM (1985) Intrinsic motivation and self-determination in human behavior. New York: Plenum.

Delgado MR (2007) Reward-Related Responses in the Human Striatum. Annals of the New York Academy of Sciences 1104:70-88.

Diekhof EK, Kaps L, Falkai P, Gruber O (2012) The role of the human ventral striatum and the medial orbitofrontal cortex in the representation of reward magnitude – An activation likelihood estimation meta-analysis of neuroimaging studies of passive reward expectancy and outcome processing. Neuropsychologia 50:1252-1266.

Dobryakova E, Jessup RK, Tricomi E (2017) Modulation of ventral striatal activity by cognitive effort. NeuroImage 147:330-338.

Duckworth AL, Peterson C, Matthews MD, Kelly DR (2007) Grit: Perseverance and passion for long-term goals. Journal of Personality and Social Psychology 92:1087-1101.

Elliot AJ, Harackiewicz JM (1996) Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. Journal of Personality and Social Psychology 70:461-475.

Gallagher S (2000) Philosophical conceptions of the self: implications for cognitive science. Trends in Cognitive Sciences 4:14-21.

Gendolla GHE, Richter M (2010) Effort Mobilization when the Self is Involved: Some Lessons from the Cardiovascular System. Review of General Psychology 14:212-226.

Göttlich, M., Beyer, F., & Krämer, U. M. (2015). BASCO: a toolbox for task-related functional connectivity. *Frontiers in Systems Neuroscience*, *9*(September), 1–10. https://doi.org/10.3389/fnsys.2015.00126

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gottlieb J, Oudeyer P-Y, Lopes M, Baranes A (2013) Information-seeking, curiosity, and attention: computational and neural mechanisms. Trends in cognitive sciences 17:585-593.

Knutson B, Taylor J, Kaufman M, Peterson R, Glover G (2005) Distributed neural representation of expected value. Journal of Neuroscience 25:4806-4812.

Krebs RM, Schott BH, Schutze H, Duzel E (2009) The novelty exploration bonus and its attentional modulation. Neuropsychologia 47:2272-2281.

Krebs RM, Boehler CN, Egner T, Woldorff MG (2011) The neural underpinnings of how reward associations can both guide and misguide attention. The Journal of neuroscience : the official journal of the Society for Neuroscience 31:9752-9759.

Kurniawan IT, Seymour B, Talmi D, Yoshida W, Chater N, Dolan RJ (2010) Choosing to Make an Effort: The Role of Striatum in Signaling Physical Effort of a Chosen Action. Journal of Neurophysiology 104:313-321.

Lee W, Reeve J (2013) Self-determined, but not non-self-determined, motivation predicts activations in the anterior insular cortex: an fMRI study of personal agency. Soc Cogn Affect Neurosci 8:538-545.

Lee W, Kim S-i (2014) Effects of Achievement Goals on Challenge Seeking and Feedback Processing: Behavioral and fMRI Evidence. PLOS ONE 9:e107254.

Lee W, Reeve J (2017) Identifying the neural substrates of intrinsic motivation during task performance. Cognitive, affective & behavioral neuroscience 17:939-953.

Liu X, Hairston J, Schrier M, Fan J (2011) Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. Neuroscience & Biobehavioral Reviews 35:1219-1236.

Loewenstein G (1999) Because It Is There: The Challenge of Mountaineering… for Utility Theory. Kyklos 52:315-343.

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets, *19*, 1233–1239. https://doi.org/10.1016/S1053-8119(03)00169-1

McClelland DC, Atkinson JW, Clark RA, Lowell EL (1976) The achievement motive: Oxford, England: Irvington.

Mobbs D, Hassabis D, Seymour B, Marchant JL, Weiskopf N, Dolan RJ, Frith CD (2009) Choking on the Money: Reward-Based Performance Decrements Are Associated With Midbrain Activity. Psychological Science 20:955-962.

Murayama K (2019) A reward-learning framework of autonomous knowledge acquisition: An integrated account of curiosity, interest, and intrinsic-extrinsic rewards. https://osfio/zey4k/.

Murayama K, Matsumoto M, Izuma K, Matsumoto K (2010a) Neural basis of the undermining effect of monetary reward on intrinsic motivation. PNAS Proceedings of the National Academy of Sciences of the United States of America 107:20911-20916.

Murayama K, Matsumoto M, Izuma K, Matsumoto K (2010b) Neural basis of the undermining effect of monetary reward on intrinsic motivation. Proceedings of the National Academy of Sciences 107:20911-20916.

Murayama K, Matsumoto M, Izuma K, Sugiura A, Ryan RM, Deci EL, Matsumoto K (2015) How Self-Determined Choice Facilitates Performance: A Key Role of the Ventromedial Prefrontal Cortex. Cerebral Cortex 25:1241-1251.

Nachev P, Kennard C, Husain M (2008) Functional role of the supplementary and pre-supplementary motor areas. Nature Reviews Neuroscience 9:856.

Nicholls JG (1984) Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. Psychological Review 91:328-346.

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Oudeyer P-Y, Baranes A, Kaplan F (2013) Intrinsically Motivated Learning of Real-World Sensorimotor Skills with Developmental Constraints. In: Intrinsically Motivated Learning in Natural and Artificial Systems (G. B, M. M, eds), pp 303-365. Berlin: Springer.

Pauli, W. M., Nili, A. N., & Tyszka, J. M. (2018). A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Scientific Data*, *5*, 1–13. https://doi.org/10.1038/sdata.2018.63

Rissman J, Gazzaley A, D'Esposito M (2004) Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage 23:752-763.

Roiser JP, Stephan KE, den Ouden HEM, Friston KJ, Joyce EM (2010) Adaptive and aberrant reward prediction signals in the human brain. NeuroImage 50:657-664.

Sedikides C, Strube MJ (1997) Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be better. Advances in Experimental Social Psychology 29:209-269.

Sinha R, Lacadie CM, Constable RT, Seo D (2016) Dynamic neural activity during stress signals resilient coping. Proceedings of the National Academy of Sciences 113:8837-8842.

Skvortsova V, Palminteri S, Pessiglione M (2014) Learning To Minimize Efforts versus Maximizing Rewards: Computational Principles and Neural Correlates. The Journal of Neuroscience 34:15621-15630.

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. Nature Neuroscience 7:887-893.

Trautwein U, Lüdtke O, Roberts BW, Schnyder I, Niggli A (2009) Different forces, same consequence: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement. Journal of Personality and Social Psychology 97:1115-1128.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Trope Y (1979) Uncertainty-reducing properties of achievement tasks. Journal of Personality and Social Psychology 37:1505-1518.

Ulrich M, Keller J, Hoenig K, Waller C, Grön G (2014) Neural correlates of experimentally induced flow experiences. NeuroImage 86:194-202.

Vostroknutov A, Tobler PN, Rustichini A (2012) Causes of social reward differences encoded in human brain. Journal of Neurophysiology 107:1403-1412.

Wisniewski D, Reverberi C, Momennejad I, Kahnt T, Haynes J-D (2015) The Role of the Parietal Cortex in the Representation of Task–Reward Associations. The Journal of Neuroscience 35:12355-12365.

Yeager DS, Walton GM, Brady ST, Akcinar EN, Paunesku D, Keane L, Kamentz D, Ritter G, Duckworth AL, Urstein R, Gomez EM, Markus HR, Cohen GL, Dweck CS (2016) Teaching a lay theory before college narrows achievement gaps at scale. Proceedings of the National Academy of Sciences 113:E3341-E3348.

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Whole brain results for parametric modulation analysis

Cluster	Cluster size	AAL region (number of voxels)	MNI peak coordinates			peak T statistic	peak z statistic
			x	y	z		
Positive relations with intrinsic motivation							
1	124	Inferior frontal gyrus, pars opercularis (R 90)	45	5	25	6.59	4.38
		Precentral gyrus (R 29)	45	14	10	4.89	3.67
		Inferior frontal gyrus, triangular (R 1)					
Negative relations with intrinsic motivation							
1	160	Fusiform gyrus (L 134)	-36	-37	-26	6.54	4.36
		Inferior temporal gyrus (L 17)	-42	-49	-17	5	3.73
		Cerebelum (L 9)	-39	-61	-20	4.71	3.59
2	261	Fusiform gyrus (R 173)	30	-52	-17	5.45	3.93
		Lingual gyrus (R 34)	30	-37	-23	5.2	3.82
		Calcarine sulcus (L 17, R 1)	33	-67	-14	4.99	3.72
		Cerebelum (R 20)					
		Parahippocampal gyrus (R 5)					
Positive relations with rewarding value for the cue							
1	167	Inferior parietal gyrus (R 65)	30	-46	46	5.57	4.09
		Supramarginal gyrus (R 32)	42	-34	43	5.31	3.97
		Postcentral gyrus (R 24)	24	-58	46	3.77	3.14
2	121	Precentral gyrus (R 91)	45	2	37	4.94	3.8
		Inferior frontal gyrus, pars opercularis (R 12)	45	2	46	4.73	3.69
		Middle frontal gyrus (R 2)	33	-1	46	4.65	3.65
		Superior frontal gyrus (R 2)					
Negative relations with rewarding value for the cue							
1	289	Fusiform gyrus (L 215)	-36	-40	-23	6.95	4.65
		Cerebelum (L 41)	-42	-64	-17	6.08	4.32
		Inferior temporal gyrus (L 21)	-36	-76	-11	4.05	3.31
		Inferior occipital gyrus (L 11)					
		Parahippocampal gyrus (L 1)					
2	365	Fusiform gyrus (R 158)	9	-85	-8	6.24	4.38
		Lingual gyrus (L 16, R 57)	-12	-100	7	5.55	4.08
		Calcarine sulcus (L 36, R 7)	27	-49	-17	5.44	4.03
		Superior occipital gyrus (L 30)					
		Middle occipital gyrus (L 25)					

Note. All *ps* < .05, cluster-level FWE corrected (with the initial thresholding at $p < .001$). If the cluster spreads over more than 5 brain regions, only the five largest contributors were named and listed.

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure captions

Figure 1. Task design and prediction. **A.** Task sequence. Participants were presented with a task cue signalling different chance of success ("Easy" = high chance of success, "Intermediate" = moderate chance of success, and "Super-advanced" is extremely low chance of success), followed by the actual task. There were also control trials ("watch-stop" trials) in which participants neither succeeded nor failed, but these trials were omitted from the figure. **B.** A schematic picture of how expected performance rewards (red line) and motivation for challenge (blue line) changes in an opposite manner as a function of task difficulty (chance of success). Expected performance reward decreases as chance of success decreases (i.e. low expected utility), whereas motivation for challenge *increases* as chance of success decreases.

Figure 2. Pattern of results as a function of group (no-reward, reward, and gambling) and chance of success (high, moderately high, and extremely-low). **A.** Self-reported intrinsic motivation for the task increased as the chance of success *decreased* in the no-reward group, whereas the pattern was opposite in the gambling group. Reward group exhibited the highest intrinsic motivation for the task with moderate chance of success. **B.** Self-reported rewarding value for the task cue showed a similar pattern except that the rewarding value for the cue monotonically decreased as the chance of success decreased in the reward group. **C.** The bilateral ventral striatum/ventral pallidum activation (right figure, *Ps* <.05, voxel-level FWE corrected) mirrored the pattern observed in the self-reported intrinsic motivation. In the no-reward group, the ventral striatum/ventral pallidum activation increased as the chance of success *decreased*.

Figure 3. Results of parametric modulation analyses in the reward group. Red shows areas that are significantly positively related to the within-person change in self-reported intrinsic

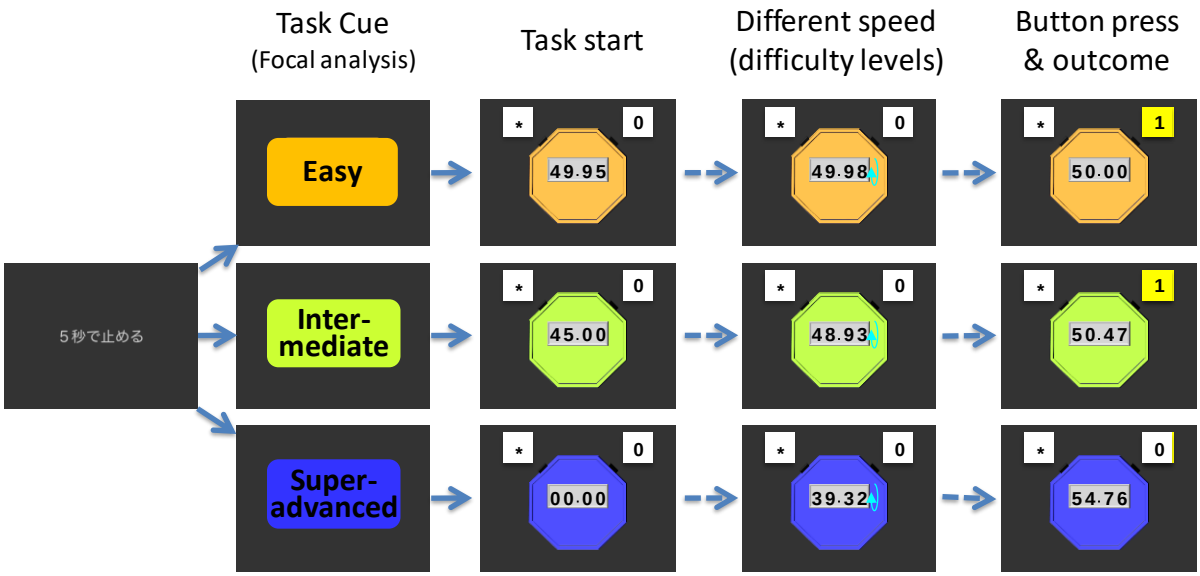
bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

motivation (i.e. activation was higher when intrinsic motivation was higher within an individual). Blue shows areas that are significantly related to change in self-reported rewarding value for the task cue (i.e. activation was higher when rewarding value for the cue was higher within an individual). All *Ps* < .05, cluster-level FWE corrected.

Figure 4. Results of functional connectivity analysis (with the ventral striatum/ventral pallidum seed) comparing reward and no-reward group for extremely-low chance of success. The results indicate functional decoupling between the left ventral striatum/ventral pallidum and the supplementary motor area when participants in the reward group expected a task with extremely-low chance of success (*P* < .05 with Bonferroni correction, cluster-level FWE corrected).

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A.



B.

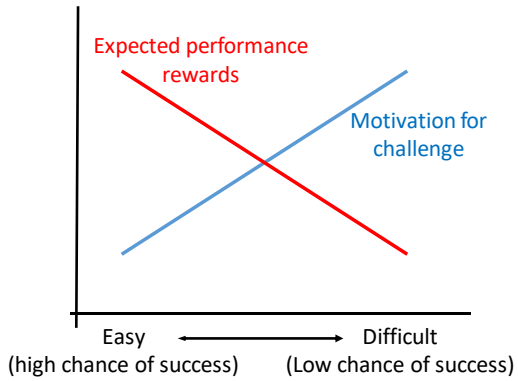


Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

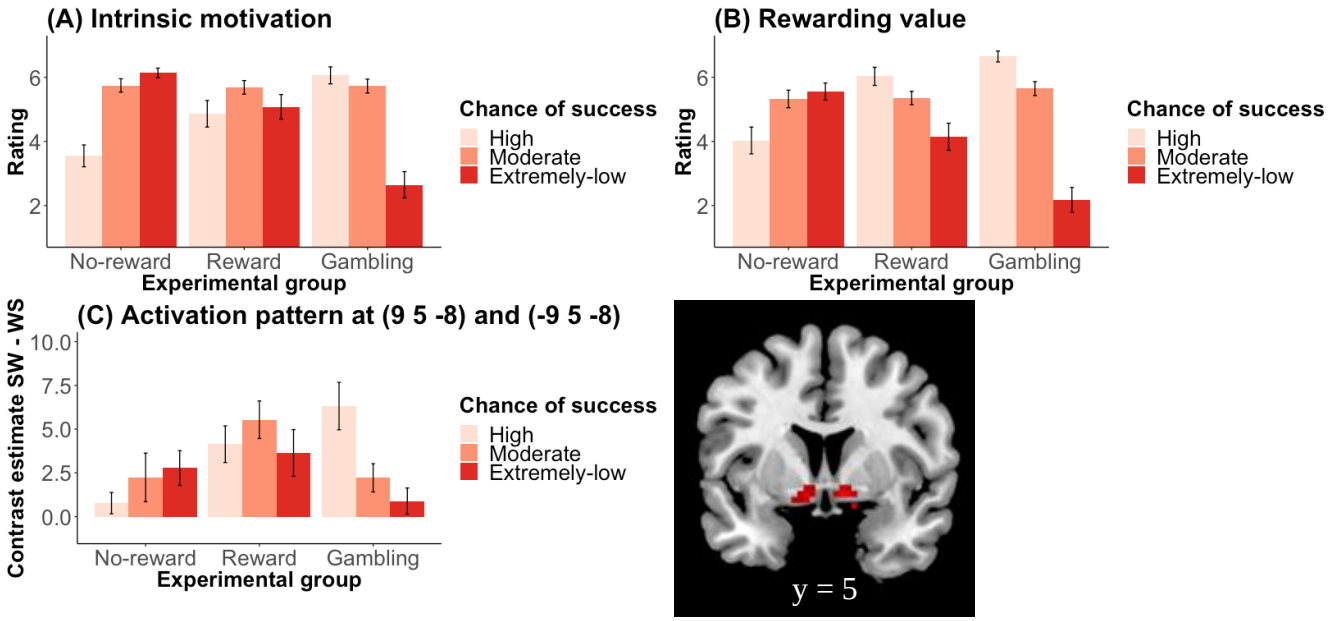


Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

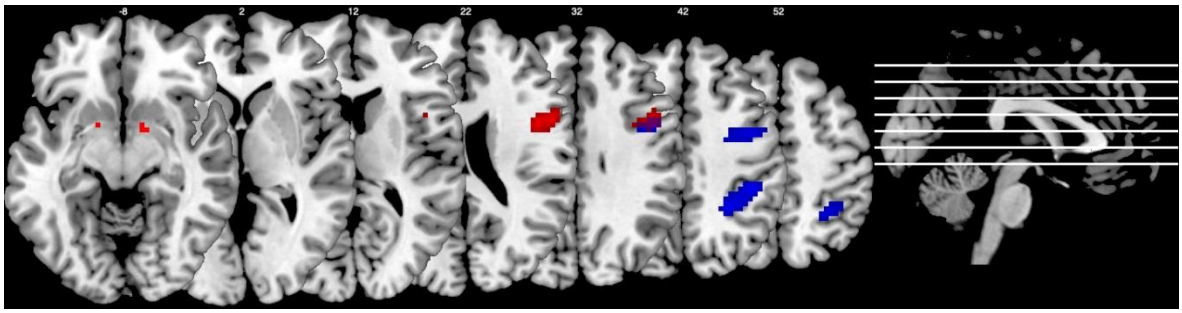


Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

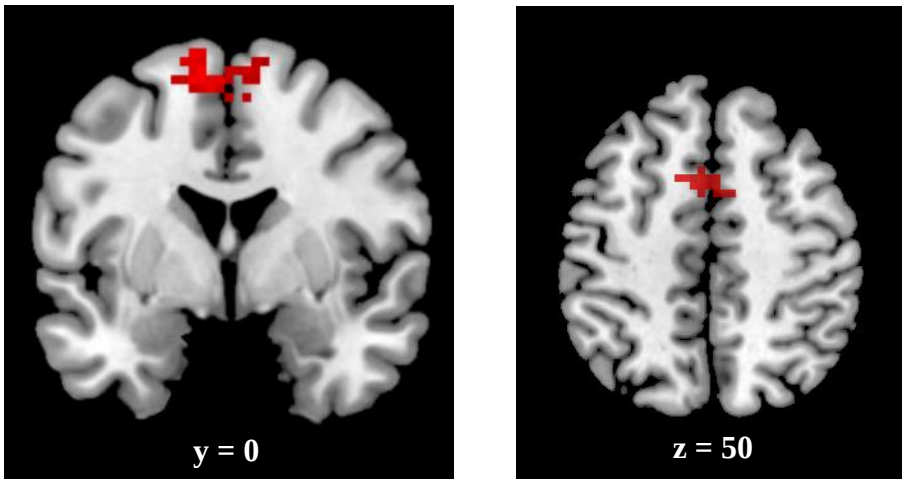


Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S1. Brain areas showing comparable activation pattern with the ventral striatum/ventral pallidum ROI.

Cluster	Cluster size	AAL region (number of voxels)	MNI peak coordinates			peak F statistic	peak z statistic
			x	y	z		
1	409	Caudate (L 36, R 29)	9	5	-11	15.42	5.62
		Thalamus (L 30, R 11)	-9	5	-8	13.16	5.16
		Putamen (L 30, R14)	-3	-7	4	9.21	4.21
		Olfactory gyrus (L 19, R 14)					
		Pallidum (L 10, R 9)					
2	4733	Lingual gyrus (L 344, R 313)	24	-91	-5	14.25	5.39
		Cerebelum (2196)	-30	-70	7	13.17	5.17
		Calcarine sulcus (L 280, R 187)	36	-52	1	11.8	4.86
		Middle occipital gyrus (L 174, R 57)					
		Inferior occipital gyrus (L 78, R 65)					
3	218	Inferior parietal gyrus (L 125)	-45	-43	49	11.09	4.69
		Postcentral gyrus (L 5)	-33	-40	34	9.01	4.15
		Angular gyrus L 3)	-39	-40	40	8.45	3.99
		Supramarginal gyrus (L 2)					
4	96	Middle frontal gyrus (R 49)	51	17	37	9.21	4.21
		Inferior frontal gyrus, pars opercularis (R 20)	33	5	40	8.47	4
		Precentral gyrus (R 14)	42	11	43	7.81	3.8
5	269	Middle frontal gyrus (L 177)	-27	11	52	8.82	4.1
		Superior frontal gyrus (L 49)	-33	26	49	8.69	4.06
		Precentral gyrus L 31)	-21	20	58	8.65	4.05
		Inferior frontal gyrus, pars opercularis (L 8)					
		Inferior frontal gyrus, pars triangularis (L 3)					

Note. All *ps* < .05, cluster-level FWE corrected (with the initial thresholding at *p* < .001). If the cluster spreads over more than 5 brain regions, only the five largest contributors were named and listed.

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S2. Results of 3 x 3 whole-brain ANOVA

Cluster	Cluster size	AAL region (number of voxels)	MNI peak coordinates			peak F statistic	peak z statistic
			x	y	z		
Main Effect of Group							
No significant activations							
Main Effect of Chance of Success							
1	1159	Lingual gyrus (L 17, R 152)	33	-52	1	23.42	5.93
		Calcarine gyrus (L 63, R 119)	-27	-70	7	18.86	5.32
		Cuneus (L 101, R 116)	12	-82	31	13.72	4.49
		Superior occipital gyrus (L 34, R 47)					
		Fusiform gyrus (R 25)					
2	135	Inferior parietal gyrus (R 74)	45	-37	46	16.13	4.9
		Supramarginal gyrus (R 31)	30	-55	37	9.89	3.73
		Angular Gyrus (R 16)					
		Postcentral gyrus (R 5)					
3	208	Cerebelum (R 148)	27	-46	-41	14.75	4.67
4	305	Cerebelum (L 197, R 10)	-12	-46	-50	14.49	4.63
		Vermis (38)	-24	-43	-47	13.4	4.44
		Fusiform gyrus (L 15)	-21	-55	-50	12.78	4.32
		Lingual gyrus (L 10)					
5	91	Lingual gyrus (L 38)	-12	-67	-11	10.96	3.96
		Cerebelum (L 53)					
Interaction Effect Group x Chance of Success							
1	3740	Cerebelum (1516)	-3	-88	-11	11.52	5.37
		Lingual gyrus (L 212, R 136)	-15	-91	-20	9.96	4.94
		Calcarine sulcus (L 170, R 41)	12	2	-11	9.63	4.85
		Middle occipital gyrus (L 152, R 57)					
		Thalamus (L 64)					
2	603	Superior frontal gyrus, medial (L 198, R 79)	0	35	58	8.69	4.56
		Supplementary motor area (L 104, R 14)	0	20	64	7.94	4.31
		Middle frontal gyrus (L 79)	-6	53	34	7.81	4.26
		Superior frontal gyrus (L 78, R 3)					
		Anterior cingulate gyrus (L 21)					
3	93	Middle temporal gyrus (L 53)	-45	-37	-11	7.09	4.01
		Inferior temporal gyrus (L 13)	-60	-46	-2	6.41	3.75

Note. All *ps* < .05, cluster-level FWE corrected (with the initial thresholding at p < .001). If the cluster spreads over more than 5 brain regions, only the five largest contributors were named and listed.

bioRxiv preprint doi: https://doi.org/10.1101/828756; this version posted November 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.