SUPPLEMENTARY METHODS: STUDY 1

Participants. Five additional subjects were recruited but excluded prior to analysis due to an imaging artifact (73 y.o. male), excessive head motion (23 y.o. male), vision trouble (71 y.o. male), medication (75 y.o. male), and inability to follow instructions (79 y.o. female). Participants were recruited from the San Francisco Bay Area and then followed up by laboratory personnel for a complete phone interview to determine eligibility. The phone interview included questions relevant to their safety and their history of physical or mental disorders (specifically stroke and neurological damage, history of heart failure, or prescription medicine shown to interfere with the blood oxygen level dependent signal, e.g., either psychiatric or cardiac). If eligible, participants completed two sessions. In the first session, participants completed a questionnaire packet, a cognitive test battery, a thorough explanation of the scanning procedures, and a practice version of the MID task. In the second session, participants engaged in the MID task while undergoing FMRI. In addition to earnings on the task, participants were paid $20/hour for their participation.

Prior to being scanned, participants received a verbal description of the task, and completed a 15-minute practice version. Participants were also shown the money that they could earn by performing the task successfully in the scanner, and all reported believing that they would receive cash based on their performance at the end of the experiment. Once in the scanner, anatomical scans were acquired. Participants then engaged in two 16-minute blocks of the incentive task and one 6-minute block of a visual localizer task during functional scan acquisition. After the scan, in addition to affective ratings, participants estimated the ratio of gain cues to lose cues (no age differences were found in ratio estimates).

Questionnaire measures. A demographics questionnaire assessed the age, marital status, current and previous occupational status, level of income, number of
years of education, and ethnicity of the participants. Several individual difference measures were included to ensure that between-group differences in self-reported affect or BOLD activation were not due to baseline group differences in trait affect or personality. The trait version of the Positive and Negative Affect Schedule (PANAS-T)\(^1\) was used to assess the extent to which participants experienced each of 22 emotional descriptors on a regular basis. A measure of physical health, the Wahler Physical Symptom Inventory (WPSI)\(^2\), asked participants to indicate how often they are bothered by each of 42 physical symptoms. The Future Time Perspective (FTP) scale\(^3\) is a 10-item self-report measure that assesses how much time people feel they have left in their lives. A 60-item short form of the Neuroticism-Extroversion/Introversion-Openness-to-Experience Personality Inventory (NEO-SF)\(^4\) asked participants to indicate their level of endorsement of each of the statements related to commonly-assessed personality traits. The 5-item Subjective Well-being and Satisfaction with Life Scale (SWLS)\(^5\) assessed general overall satisfaction with life.

**Neuropsychological battery.** The Mini-Mental Status Exam (MMSE)\(^6\) was administered to all participants as a screen for dementia. Three subtests from the Wechsler Adult Intelligence Scale Third Edition (WAIS-III)\(^7\) with well-validated ranges for older adults were administered to each participant. The WAIS-III Digit Span test requires that participants repeat numerical strings forward and backward. It is considered a measure of working memory and correlates well with general intelligence. The WAIS-III Digit Symbol test requires participants to match symbols with letters as quickly and accurately as possible in a 120-second period. The WAIS-III Vocabulary test requires that participants provide definitions for words presented in both written and spoken form, and correlates well with verbal intelligence. Two subtests, Verbal and Category Fluency, of the Delis-Kaplan Executive Function System\(^8\) were administered. The Verbal Fluency (FAS) subtest requires that participants name as many words as possible beginning with a given letter (first F, then A, then S) in a 60-second period. The
similar Category Fluency subtest requires that participants name as many words as possible that fall into the given category (animals) in a 60-second period. The Trail Making Test (TMT) from the Halstead-Reitan Neuropsychological Test Battery has two parts (A & B) which are both timed until completion. The first part (Trails A) requires that participants sequentially connect 25 encircled numbers on a standard sheet of paper. The second part (Trails B) requires that participants connect a series of numbers and letters in an alternating pattern. Trails B is considered to be a good indicator of general frontal lobe cognitive function.

**Visual localizer task.** Collection of FMRI data in older adults raises many methodological issues, which necessitate careful sampling and measurement. Even assuming good health, the hemodynamic response of older individuals has been shown to be similar but more variable than that of younger adults in cortical regions. Thus, a visual localizer task was included to examine potential age differences in individual hemodynamic response functions (HRFs). The task consisted simply of responding with a button press to flickering checkerboard stimuli that were presented for 2 s, separated by random interstimulus intervals ranging from 2–38 s. Timecourses of activation were extracted from voxels in primary visual cortex (V1) in individual participants. A multivariate GLM revealed no significant effect of age ($F_{1, 11} = 1.214, P = .371$). Additionally, it should be noted that none of the participants included in this study have abnormally shaped HRFs (Supp. Fig. 15) as has been found previously in FMRI studies of older adults.

**VOI Definition.** VOI spheres were manually adjusted for individual participants to account for potential anatomical variability between the age groups not corrected for by the Talairach warping procedure and in order to avoid partial voluming of functional signal. The definition procedure began with *a priori* coordinates selected from previous data sets which could be shifted in two dimensions within a 10 mm x 10 mm constrained region along at least one fixed plane. An algorithm was created for each
VOI. For the VS, the start coordinates were 11, 12, 0 with the coronal and axial planes fixed. If imposing a 6 mm diameter sphere resulted in sampling of the neighboring ventricle, the sphere was shifted within 10mm right/left. However, for the VS, the a priori coordinates were compatible with the anatomy of nearly participant (only two participants required shifting to the right). For the MCAUD, the start coordinates were –9, 13, 9 with the coronal plane fixed. If imposing a 6 mm diameter sphere resulted in sampling of the neighboring ventricle, the sphere was shifted within 10mm right/left. For the MCAUD, over half of the participants required shifting to the left and only one participant requiring a 1mm shift inferior. For the AINS, the start coordinates were 39, 19, 7 with the sagittal plane fixed. If imposing a 6 mm sphere resulted in sampling of the neighboring CSF, the sphere could be shifted within 10 mm right/left or superior/inferior. All but 4 participants required at least a 1 mm shift in at least one plane. For the MPFC, the start coordinates were –1, 53, –6 with the coronal plane fixed. If imposing a 6 mm sphere resulted in sampling of the neighboring CSF, the sphere could be shifted within 10 mm right/left or superior/inferior. All 24 participants needed at least a 1 mm shift in at least one plane (Supp. Table 5).
SUPPLEMENTARY METHODS: STUDY 2

Participants. Twelve younger (age 19–34, six female) and twelve older (age 65–81, six female) adults were recruited from the San Francisco Bay Area and completed one session. Seven younger and all twelve older adults from Study 1 participated in Study 2. An additional five younger adults were new but matched the younger adults from Study 1 on age, education, and socioeconomic status. All participants gave written informed consent, and the experiment was approved by the Institutional Review Board of the Stanford University Medical School. Each participant completed a demographic questionnaire, received a verbal description of the task, and completed a 30 trial practice version. Participants were also shown the money that they could earn by performing the task successfully. Participants then engaged in two 120 trial runs of an incentive learning task. In addition to earnings on the task, participants were paid $20/hour for their participation.

Monetary incentive learning (MIL) task. The design of the monetary incentive learning (MIL) task was inspired by a similar recently published incentive learning paradigm \(^1\). Across both runs, the entire task included 240 trials. During each trial, participants chose from a pair of abstract cues (decision), viewed their highlighted choice on screen, and received feedback about how much money they won or lost on the trial (outcome) (Supp. Fig. 17). The display duration of the first frame of the task was self-paced to accommodate differences in vision and decision reaction time among younger and older participants. Three pairs of cues were used in each run (gain, loss avoidance, neutral). Different pairs of cues were used during practice, run 1, and run 2 to avoid age-related impairments in reversal learning. Within gain and loss avoidance pairs one cue yielded a high probability optimal outcome (60% +$1.00, 40% +$0.00; 60% −$0.00, 40% −$1.00) and the other a low probability optimal outcome (30% +$1.00, 70% +$0.00; 30% −$0.00, 70% −$1.00). Both cues always had no impact on winnings (100% $0.00) in the neutral condition. Each cue within each pair appeared
equally often on the left and right side of the screen within runs. The pairing of cues with outcomes was counterbalanced across participants. The goal of the experiment was to learn which cue in each pair was higher in expected value (high probability gain acquisition, high probability loss avoidance). Each of three trial types was presented 40 times per run in an individually randomized order for each participant.

Hits were calculated as the percentage of correct responses per condition (i.e., the cue associated with a higher expected value) and averaged between runs. As the goal of this study was to test for a significant impairment in loss avoidance but not gain acquisition among older adults, hit rate was analyzed with mixed-model ANOVAs with incentive valence (gain acquisition, loss avoidance) and trial quarter (first 10 trials, second 10 trials, third 10 trials, last 10 trials) as the within-subject factors and age (younger, older) as the between-subject factor. Post-hoc analyses compared hits across all conditions (gain acquisition, loss avoidance, neutral) for each group with within-subject t-tests versus chance (50%) (corrected for six comparisons, $P < .008$). Additional post-hoc tests assessed learning differences between groups over time by comparing accuracy within all four quarters (ten trials per quarter) across both incentive conditions (gain acquisition, loss avoidance) with between-subject t-tests (corrected for eight comparisons, $P < .006$).
References


