

# Altered behavioral and neural responsiveness to counterfactual gains in the elderly

Michael J. Tobia<sup>1,2</sup> · Rong Guo<sup>1,3,4</sup> · Jan Gläscher<sup>1</sup> · Ulrike Schwarze<sup>1</sup> · Stefanie Brassens<sup>1</sup> · Christian Büchel<sup>1</sup> · Klaus Obermayer<sup>3</sup> · Tobias Sommer<sup>1</sup>

Published online: 10 February 2016  
© Psychonomic Society, Inc. 2016

**Abstract** Counterfactual information processing refers to the consideration of events that did not occur in comparison to those actually experienced, in order to determine optimal actions, and can be formulated as computational learning signals, referred to as fictive prediction errors. Decision making and the neural circuitry for counterfactual processing are altered in healthy elderly adults. This experiment investigated age differences in neural systems for decision making with knowledge of counterfactual outcomes. Two groups of healthy adult participants, young ( $N = 30$ ; ages 19–30 years) and elderly ( $N = 19$ ; ages 65–80 years), were scanned with fMRI during 240 trials of a strategic sequential investment task in which a particular strategy of differentially weighting counterfactual gains and losses during valuation is associated with more optimal performance. Elderly participants earned significantly less than young adults, differently weighted counterfactual consequences and exploited task knowledge, and exhibited altered activity in a fronto-striatal circuit while

making choices, compared to young adults. The degree to which task knowledge was exploited was positively correlated with modulation of neural activity by expected value in the vmPFC for young adults, but not in the elderly. These findings demonstrate that elderly participants' poor task performance may be related to different counterfactual processing.

**Keywords** Fictive prediction error · Model-based fMRI · Ventromedial PFC · Reward prediction · Aging

Counterfactual information processing refers to the consideration of events that did not occur in comparison to those actually experienced in order to determine optimal actions. Counterfactual information processing is a ubiquitous component of cognition (Byrne, 2002) and an important source of bias during decision making (Coricelli & Rustichini, 2010; Lohrenz, McCabe, Camerer, & Montague, 2007; Montague, King-Cassas, & Cohen, 2006). It can be conceptualized as an episodic memory function (Barbey, Kruger, & Grafman, 2009; Byrne, 2002; De Brigard, Addis, Ford, Schacter, & Giovanello, 2013; Horhota, Mienaltowski, & Blanchard-Fields, 2012), or formulated as fictive prediction errors (FPE) in computational models of reward-based learning and decision making (Lohrenz et al., 2007; Montague et al., 2006; Tobia et al., 2014). An FPE is similar to a standard reward prediction error in that it is a quantity that adjusts future expected reward values but differs in that it is computed for actions that were not selected with rewards that were missed. Thus, in contrast to standard reward prediction errors, FPEs allow the acquisition of expected values even for non-selected actions. It has been shown that young adults learn to predict rewards more quickly and accurately when they incorporate counterfactual outcomes (Tobia et al., 2014), but whether learning to predict rewards from counterfactual

---

**Electronic supplementary material** The online version of this article (doi:10.3758/s13415-016-0406-7) contains supplementary material, which is available to authorized users.

---

✉ Michael J. Tobia  
mtobia@hmc.psu.edu

- <sup>1</sup> Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
- <sup>2</sup> Center for NMR Research, Department of Radiology, Pennsylvania State University College of Medicine, Hershey, PA, USA
- <sup>3</sup> Department of Software Engineering and Theoretical Computer Science, School IV Electrical Engineering and Computer Science, Technische Universität Berlin, Berlin, Germany
- <sup>4</sup> Bernstein Center for Computational Neuroscience, Berlin, Germany

consequences is affected by age has not been investigated. This experiment employed a model-based fMRI experimental design to investigate whether age affects learning from counterfactual outcomes, and if differences are reflected in the underlying neural activity of a fronto-striatal circuit previously associated with reward prediction, prediction error, and counterfactual outcome processing (Brassen, Gamer, Peters, Gluth, & Büchel, 2012; Büchel, Brassen, Yacubian, Kalisch, & Sommer, 2011; Gläscher, Hampton, & O'Doherty, 2009; Hare et al, 2008; Lohrenz et al, 2007; McClure et al., 2004; O'Doherty, 2004; Tobia et al., 2014).

The model-based fMRI design operationally defines certain cognitive processes as a set of computational model parameters that vary trial by trial, which can then be used to predict neural responses and localize the computational parameter to functional anatomy (Borst & Anderson, 2013; Gläscher & O'Doherty, 2010). It has been successfully applied to decision-making research and uncovered a circuit between the ventral striatum and ventromedial prefrontal cortex (vmPFC), which together form a system for decision making based on learned action-specific expected values (Hare et al., 2008). From model-based fMRI research, the vmPFC is postulated to process representations of expected value that correspond to choice behavior (Gläscher et al., 2009), and the striatum is postulated to process prediction errors that subsequently modify expected values (McClure et al, 2004; O'Doherty, 2004), thereby adapting both brain and behavior over time.

The fronto-striatal circuit of reward-based learning differs between young and older adults (Grady, 2012; Hedden & Gabrieli, 2004; Samanez-Larkin & Knutson, 2015). Older adults are characterized by decreased gray matter and white matter (WM) volume in voxel-based morphometry studies (Brickman, Habeck, Zarahn, Flynn, & Stern, 2006) as well as reduced measures of WM integrity from diffusion weighted imaging (Bennet, Madden, Vaidya, Howard, & Howard, 2010; Burzynska et al., 2010; Pfefferbaum, Adalsteinsson, & Sullivan, 2005; Samanez-Larkin, Levens, Perry, Dougherty & Knutson, 2012). In some cases, such as in multialternative choice tasks, age-related changes in structure and functioning of the fronto-striatal circuit are correlated with age-altered decision making and performance (Grady, 2012; Hedden & Gabrieli, 2004; Samanez-Larkin & Knutson, 2015) as well as with different neural responsiveness to reward, reward predictions, and reward prediction errors (Eppinger, Nystrom, & Cohen, 2012; Eppinger, Walter, Heekeren, & Li, 2013; Eppinger, Schuck, Nystrom, & Cohen, 2013; Samanez-Larkin, Wagner & Knutson, 2011; Samanez-Larkin et al., 2007; Samanez-Larkin, Kuhnen, Yoo, & Knutson, 2010; Samanez-Larkin, Worthy, Mata, McClure, & Knutson, 2014; Vink, Kleerekooper, van den Wildenberg, & Kahn, 2015). But the relation between age-related changes in structure and functioning of the brain, and age-related differences in task

performance, is not always consistent. For example, one study reported that older adults make more decision-making mistakes related to risk taking, which was attributable to greater variability in the ventral striatum compared to young adults (Samanez-Larkin et al., 2010). Another study of decision making in healthy aging demonstrated that older adults show reduced activity in the vmPFC during choice behavior, but this age difference in neural activity was uncorrelated with behavior, which was suboptimal (Eppinger, Heekeren & Li, 2015). This suggests that task performance irrespective of age mediates the relationship between neural activity and decision making based on reward predictions. Indeed, vmPFC activation is correlated with expected values, as it is in young adults, for older adults who perform more optimally (Samanez-Larkin et al., 2011). Moreover, advanced age is associated with different decision-making strategies depending on the task contingencies (Worthy & Maddox, 2012) and older adults choice behavior is sometimes better explained by a win-stay-lose-shift (WSLS) strategy that does not rely on a reward prediction (Worthy & Maddox, 2012), as opposed to reinforcement learning strategies that do. This supports the notion that although the elderly use different decision-making strategies dependent on the task context, the vmPFC is important for processing reward predictions to guide choice behavior in some contexts. Taken together, these data suggest that older age groups' relatively poor decision-making task performance is related to the decline in the fronto-striatal circuit of reward-based learning, which results in altered subjective values to guide choices (Eppinger et al., 2015; Samanez-Larkin et al., 2011).

Reward predictions from the fronto-striatal circuit can be affected by counterfactual information processing (Brassen, Gamer, Peters, Gluth, & Büchel, 2012; Gu, Kirk, Lohrenz & Montague, 2013; Lohrenz et al., 2007; Tobia et al., 2014), and there is evidence that fictive prediction error (FPE) processing overlaps reward prediction error processing in the ventral striatum (Büchel et al., 2011; Tobia et al., 2014). Furthermore, counterfactual information processing is supported by the medial PFC and lesions to this region reduce the influence of counterfactual comparison on decision making (Barbey et al., 2009; Levens, Larsen, Bruss, Tranel, Bechara, & Mellers, 2014; Sommer, Peters, Gläscher & Büchel, 2009). As such, age-related differences in this brain system may lead to altered counterfactual information processing and decision-making strategies in older adults. Indeed, there is evidence that healthy senior adults generate less episodic counterfactual scenarios during laboratory tasks (Horhota et al., 2012), and they are less sensitive to counterfactual feedback about both optimal and suboptimal behavior (Brassen et al., 2012). Together, this suggests that knowledge from counterfactual information may not be effectively incorporated into reward predictions and subjective value by the elderly during decision making.

To answer this question, we investigated whether there are age-related differences in decision making based on reward predictions learned with counterfactual outcomes, and tested if behavioral differences reflect age-altered functioning of the underlying neural system using a model-based fMRI experimental design. The strategic sequential investment task (SSIT; Tobia et al., 2014) was designed to investigate the effects of counterfactual consequences on action-specific value learning within a modified Q-learning computational framework (Watkins & Dayan, 1992) and a model-based fMRI experimental design (Borst & Anderson, 2013; Gläscher & O'Doherty, 2010). During the task, there is an event dedicated to presenting counterfactual outcome information explicitly to the participants, similar to the presentation of a reward from which reward prediction errors can be computed. Thus, the SSIT encourages participants to utilize counterfactual rewards when learning to decide on the basis of reward predictions. Previous research utilizing the SSIT in a model-based fMRI experimental design (Tobia et al., 2014) demonstrated that young adults who successfully exploited their knowledge (i.e., accrued more reward) weighted counterfactual gains less, and weighted counterfactual losses more, than poorly performing participants. The counterfactual Q-learning model (FPEQ) parameters for expected value and prediction error were correlated with neural activity in the vmPFC and striatum, respectively, which is consistent with previous model-based fMRI research (Gläscher et al., 2009; Hare et al., 2008; McClure et al., 2004; O'Doherty, 2004). These findings show that the SSIT-FPEQ model-based fMRI paradigm is sensitive to the effects of disparate counterfactual valuation strategies, and it relates to the fronto-striatal circuit previously implicated in reward prediction and prediction error processing. Thus, the SSIT-FPEQ model-based fMRI paradigm is well-suited to investigate age group differences in learning from standard reward prediction errors and FPEs, and neural activity of the fronto-striatal value learning system.

## Method

### Participants

Nineteen healthy elderly adults, ages 65 to 80 years, were recruited from the community to participate in the experiment. They were carefully screened by a clinical neuropsychologist for mental health impairments using German tests for dementia and fluid intelligence (CERAD and the Leistungsprüfsystem Subtest 3 and 4). Initial screening of participants was done with the German version of the *Consortium to Establish a Registry on Alzheimer's Disease–Neuropsychological Assessment Battery* (CERAD-NAB; Thalmann et al., 2000). The total score was computed according to Chandler et al. (2005), and

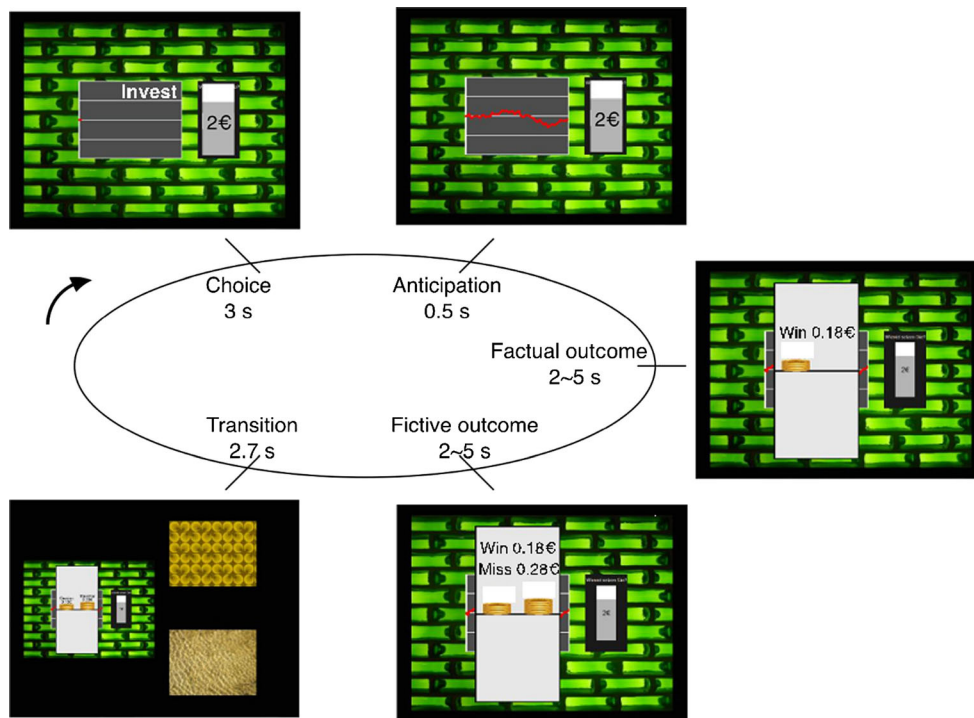
only participants were included that scored above Chandler's cut-off score of 85.11 (our sample:  $M = 87.9$ ,  $SD = 4.8$ , range 86–94). In addition, participants were screened for fluid intelligence using Subtests 3 and 4 of the German intelligence test "Leistungsprüfsystem" with a cut-off of  $z < -1.0$  (LPS; Kreuzpointner, Lukesch, & Horn, 2013). The mean  $z$  scores of our sample was for LPS 3 = 0.71 ( $SD = 0.9$ , range -0.7–2.1) and for LPS 4, mean = 1.1,  $SD = 0.5$ , range 0.4–1.8). Individuals meeting criteria for depression were excluded, using the German version of the BDI (Hautzinger et al., 1995), and a cutoff of 11 (our sample mean 3.3,  $SD = 2.7$ , range 0–10). The final sample consisted of 19 elderly subjects the mean age 66.4 ( $SD = 5.0$ , range 61–78, 13 male).

Data from 30 healthy young adult males ( $M = 23.7$  years,  $SD = 2.9$ , range 18–30 years) from a previous study (placebo group of Tobia et al., 2014) were used for age comparisons in this study. In the previous study we used a dietary depletion design to reduce dopamine and serotonin levels. The depletion might affect transmitter levels differentially in males and females, which could add noise to the data. Therefore, we included only male participants. In the current study, we included females in the elderly group because we did not expect any differential effects of counterfactual processing between sexes. All protocols were approved by the ethics committee of the medical association of Hamburg, and carried out in accord with the Declaration of Helsinki.

### Strategic sequential investment task (SSIT)

The SSIT is a reward-based value learning and intertemporal choice paradigm that involves a hierarchical state space for sequential investment choices and presents both factual and counterfactual outcomes separately on each trial. Screenshots of an example trial are shown in Fig. 1. The choice event is presented with a state-specific neutral background image and is followed by a brief anticipation phase (combined 3,500 ms duration). Participants enter their choice during this event. Then, the outcome of the trial is presented indicating the amount gained or lost by a stack of coins (2,000–5,000 ms duration). Counterfactual outcomes are then presented as a separate event (2,000–5,000 ms duration) with a stack of coins indicating how much more could have been won or lost with a maximal investment. Last, participants are shown the state transition (2,700 ms duration) selected by their investment magnitude, as well as the state transition that did not happen.

Figure 2 displays the task state space. The seven states of the SSIT indicate fictive stocks (indicated by uniquely identifiable background stimulus) that differ with respect to their winning probabilities, as well as to the magnitude of potential wins and losses (i.e., four states have negative expected values, and three states have positive expected values). A round of trials consisted of three decisions, one in each of three different states, which comprised 1 of 4 possible paths

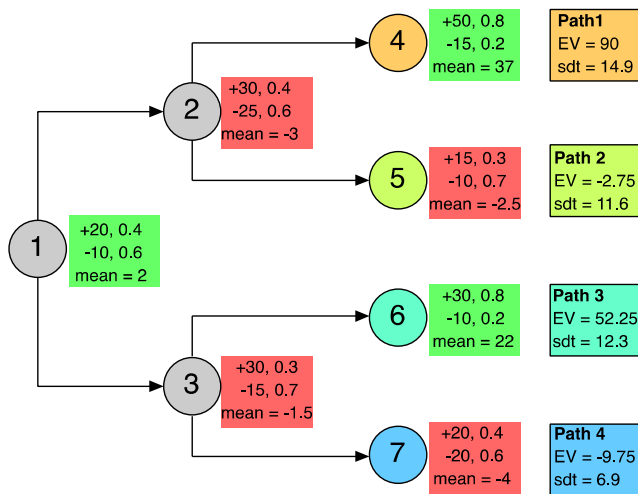


**Fig. 1** Screenshots from an example SSIT trial. The task presents five separate events in sequence, starting with the upper left and proceeding clockwise. The first event is the choice event, during which time the participants enter their wagers. The second event is a brief anticipation phase prior to the outcome. The third event shows the outcome of the trial

and the amount won or lost. The fourth event shows the counterfactual outcome (i.e., how much more could have been won or lost with a maximal wager). The final event shows the state transition, and a new trial begins after a short intertrial interval

through the state space. Each round of trials begins in State 1 and ends in either State 4, 5, 6, or 7. Participants choose in each trial to gamble 0€, 1€, 2€, or 3€ that a fictive stock will increase in value. The outcome of each trial (i.e., increase or decrease in a fictive stock value) is probabilistic, and losses occur more frequently than wins in most states. Because the

probability of winning and the payoff structure for each state produces the same outcome regardless of the selected wager, the effect of the wager is to scale the amount gained or lost. For example, investments of 1€ or 3€ differ only in the amount that each would return as a gain or loss on a given trial. As such, larger investments (2€–3€) imply greater risk to lose more and are referred to as high risk (HR), and lower investments (0€–1€) are referred to as low risk (LR).



**Fig. 2** The state space for the SSIT. The box next to each state number displays the expected win/loss magnitude (top/bottom) for each trial and long-term value of each state indicated by the mean expected value. State 4 has the highest mean expected value, and the path that ends in State 4 has the highest expected value of all paths

Choices directly control state transitions deterministically, but the behavioral contingencies that control the state transitions are unknown to the participants in advance of the task. If the participant makes an LR gamble, then the next state transition will be into a nonlucrative path. If the participant makes an HR gamble, then the next state transition will be into a lucrative path. Participants must learn to make choices that accept moderate interim losses in order to steer state transitions toward a lucrative goal state and optimize long-term gains (i.e., intertemporal choices). For example, in order to access the most lucrative State 4 (highest probability of winning and highest payoff), participants need to risk losing moderate amounts in States 1 and 2 by selecting HR gambles of 2 €–3€. Selecting an LR gamble (0€–1€) transitions to a less lucrative path, and precludes access to State 4 and optimal earnings. The optimal strategy to maximize earnings (and minimize losses) in the SSIT is to select an investment pattern of 2€, 2€, 3€, in states 1, 2, and 4, respectively.



Counterfactual consequences are instantiated in the SSIT as monetary gains and losses that were not experienced because they are contingent on forgone, alternative choices of the participant in a given trial. Such counterfactual gains and losses (also referred to as fictive outcomes) are examples of downward and upward counterfactual comparisons, respectively. A counterfactual gain (downward comparison) occurs when an alternative action (i.e., investment) would have led to a greater loss, for which the counterfactual outcome is experienced as a gain (savings) on a losing trial (e.g., investing everything instead of a smaller portion would have led to an even greater loss). A counterfactual loss (upward comparison) is a missed opportunity that occurs when an alternative action would have led to an increased gain, for which the counterfactual outcome is experienced as a loss (cost) on a winning trial (e.g., investing everything instead of a smaller portion would have led to an even greater gain). Such fictive reward signals can be used in the SSIT to further inform valuation beyond the information provided by standard reward prediction errors alone, leading to more rapid acquisition of optimal action values. The counterfactual consequences clue participants about the potentially high expected value of certain patterns of sequential choices despite frequent losses (i.e., Paths 1 and 3). Specifically, attending to counterfactual losses biases choices toward outcomes that are optimal in the long run (i.e., Paths 1 and 3), whereas attending to counterfactual gains could bias choices toward outcomes that are severely nonoptimal in the long run (Paths 2 and 4). For example, adjusting behavior in response to counterfactual losses guides future choices toward larger investments (i.e., “I could have won more, so I will invest more”) and more optimal gains over the long run because the participant ends up in the very lucrative State 4. In contrast, adjusting behavior in response to counterfactual gains biases decisions toward smaller investments (i.e., “I could have lost more, so I will invest less”) and therefore suboptimal SSIT performance.

At the end of each set of 10 rounds of the SSIT (30 trials), participants rated their experienced win frequency for each of the seven SSIT decision states (cued by their respective visual patterns) on a scale from 0 = *never win* to 5 = *always win* (fMRI data were not recorded during these ratings). These ratings were collected eight times in total throughout the task and were used to confirm that participants were engaged with the task and able to acquire knowledge about the individual states but did not assess knowledge of the overall task structure.

Prior to entering the MRI scanner for data collection, all participants were given instructions of how to manipulate the mouse to make choices and completed 15 to 20 practice trials in order to become familiar with the aims of their task. Participants received an explanation that the task contained several conditions, that each condition would be related to a different colored pattern, and that each colored pattern was associated with a different probability of winning or losing.

They were shown that they start with an account value of 0€, and were instructed that the goal of the task was to figure out how to make choices so that they could earn as much money as possible to fill their account. Practice trials did not contain details of the actual SSIT task contingencies or payoff structure used in the experiment, and participants were not informed that state transitions could be controlled by their choices. All participants confirmed they understood how to manipulate the mouse, that their goal for the task was to earn as much money as possible, and that each trial involved an opportunity to win as well as a risk of losing money.

### Computational modeling

The *Q-learning* model (Watkins & Dayan, 1992) estimates the *Q-value* (i.e., expected value) of a chosen *action* given the observed state in an online fashion. The probability of a particular choice is nonlinearly related to the *expected sum of discounted future rewards* (i.e., expected value), which has to be estimated from observations as with a temporal difference (TD) reward prediction error signal. However, the measurable TD error incorporates only factual consequences stemming from the chosen action. To model the decisions of the participant in the SSIT task we therefore extended *Q-learning* with counterfactual consequences in a two-stage update process.

We developed a computational model of human SSIT performance (FPEQ), which nests the standard *Q-learning* (StdQ) algorithm (Watkins & Dayan, 1992). FPEQ incorporates fictive rewards by updating action-specific expected values with fictive prediction error (FPE) signals in a second stage of processing on each trial, so that the *Q-values* are adjusted twice on each trial. The first update is from the standard *Q-learning* temporal difference (TD) update, and the second corresponds to the FPE update (explained below). Counterfactual gains ( $f^-$ ) and losses ( $f^+$ ) are incorporated into separate FPEs to dissociate their contributions to valuation. After applying the standard TD reward prediction error update to the chosen action, the FPEQ model computes an FPE- or FPE+ from the  $f^-$  and  $f^+$  associated with each forgone action and updates their action values accordingly. Counterfactual gains,  $f^-$ , and counterfactual losses,  $f^+$ , specifically refer to the fictive gain or loss associated with each possible counterfactual outcome for each trial. These are used to compute fictive prediction errors (FPEs) within the model framework and are not interchangeable because FPEs represent quantities computed by the model. Previous research utilizing the SSIT in a model-based fMRI experimental design with young adults (Tobia et al., 2014) demonstrated superior goodness of fit for FPEQ compared to StdQ, which is directly attributable to the incorporation of FPEs as action-specific value learning signals because the FPEQ model nested StdQ. The SSIT task consists

of seven states each indicated by a unique background stimulus, where subjects choose among four actions of investing 0€, 1€, 2€, or 3€. We denote the market change of a stock by  $o_t$ , which is drawn from the reward function of each state as described in Fig. 2. We denote the action at trial  $t$  as

$a_t := \{0, 1, 2, 3\}$  in the state  $s_t := \{1, 2, \dots, 7\}$ . For the Q-learning model, all Q-values were initialized with 0. After choosing action  $a_t$  in the current state  $s_t$ , observing the successive state  $s_{t+1}$  and received reward  $r_t$ , standard Q-learning updates only the Q-value of the current state-action pair:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha \underbrace{[ r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) ]}_{\text{TD-error}} .$$

The *learning-rate* parameter  $\alpha$  determines the pace of changes in behavior and *discount factor*  $\gamma$  reflects the preference of short-term over long-term rewards. Note that standard Q-learning (StdQ) ignores the counterfactual information provided in SSIT.

FPEQ modeled the effects of counterfactual information in a second TD-like update stage, in which the subject is presented with the counterfactual outcome associated with having invested the maximum amount. Analogously to Lohrenz et al. (2007) and Chiu, Lohrenz, and Montague (2008), we defined the experienced *fictive prediction error* (FPE) differently for positive and negative market outcomes:

- When the market goes up and less than 3€ is invested, the *counterfactual loss* is defined as the amount of money one could have won more (i.e.,  $f_+ := 3o_t - r_t$ ).
- When the market goes down and more than 0€ is invested, the *counterfactual gain* is defined as the amount of money one would have lost more (i.e.,  $f_- := r_t - 3o_t$ ).

For example, if the subject experienced a positive stock value change of 10% after betting 1€, the counterfactual loss is thus  $f_+ := (3€ - 1€) \cdot 10\% = 0.2€$ . For a negative stock value change, on the other hand, the counterfactual gain would be  $f_- := (1€ - 3€) \cdot (-10\%) = 0.2€$  as well. The counterfactual information modifies the Q-learning algorithm in a second update phase within the same trial by using the fictive reward signal ( $f_+$  or  $f_-$ ) as input to a TD computation for the more optimal unselected actions. In contrast to StdQ, which does not update any action's Q-value with FPEs, *all* actions that would have yielded better reward (unchosen action) are updated by FPEQ, and the chosen action is similarly modified; that is all  $a \geq a_t$  if market goes up and all  $a \leq a_t$  if market goes down. For each of those actions, the  $f := \{f_+, f_-\}$  takes over the role of reward, and the influence of counterfactual losses and gains is modeled independently by different learning rates  $\alpha_{FPE} := \{\alpha_+, \alpha_-\}$ :

$$Q(s_t, a) = Q(s_t, a) + \alpha_{FPE} \underbrace{[ f + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a) ]}_{\text{fictive TD-error}} .$$

The model assumes that the participants learn over time to which of the two possible next states (e.g., States 2 and 3 after State 1) they will be transferred dependent on their decision. With this formulation, both fictive error signals (counterfactual gains and losses) produce congruent effects and always increase the Q-value of the associated action. Thus, *unchosen* actions that would have produced a better outcome than the selected action increase in value on each trial. But this produces differential effects on the *chosen* action depending on whether the trial outcome was factually winning (i.e., increase in stock value) or factually losing (i.e., decrease in stock value) because the Q-value of the chosen action is first modified by the gain or loss with the TD, and then subsequently increased by the FPE. As such, on winning trials, the Q-value for the

chosen action is increased twice, and on losing trials the Q-value is first decreased and then adjusted in the opposite direction.

Both the StdQ and the FPEQ model determine the likelihood of a chosen action with a softmax:

$$P(s_t, a_t) = \frac{\exp(\beta \cdot Q(s_t, a_t))}{\sum_{a'=0}^3 \exp(\beta \cdot Q(s_t, a'))} .$$

The FPEQ model contains five parameters: the discount factor  $\gamma$ , the inverse temperature  $\beta$ , the factual learning rate  $\alpha$ , and the counterfactual loss/gain learning rates  $\alpha_+$  and  $\alpha_-$ . Standard Q-learning is a special case (i.e., nested) with  $\alpha_+ = 0$  and  $\alpha_- = 0$ . All the free parameters from both models were

individually fitted to each subject by minimizing the negative log likelihood of the data given different parameter settings using MATLAB's `fminunc` function. The goodness of fit for StdQ and FPEQ models were compared using likelihood ratio test and BIC scores. Direct comparison of model parameters between age groups was performed using a two-sample *t* test ( $df = 47$ , two-tailed significance).

### Additional computational models

In addition to the StdQ and FPEQ models described above, we also examined the fit of a model-based-Q (Doll, Simon, & Daw, 2012; Gläscher, Daw, Dayan, & O'Doherty, 2010) and win-stay-lose-shift Q (Worthy & Maddox, 2014) model that could potentially predict the SSIT choice behavior of either younger or older adults better than our counterfactual-based FPEQ model. Worthy and Maddox showed that older adults' choice behavior was better explained by the WSLS model when the outcomes of sequential choices are independent of each other, but utilized reward predictions to guide choice behavior when the outcomes of sequential choices were dependent. The SSIT is a choice-dependent task for which the outcome of a series of decisions is affected by current decisions. As such, we expected that a model with reward predictions would be a better fit to both groups' behavior. The model-based Q (Gläscher et al., 2010; Doll et al., 2012) postulates that participants make decisions based on a learned cognitive model of probabilistic state transitions, similar to a spatial navigation task, rather than learning state-specific reward predictions per se as in cache-based (i.e., model-free) decision making. Previous research has shown that some task performance is accounted for by a model-based decision-making mechanism that predicts probabilistic state transitions, rather than rewards per se (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Gläscher et al., 2010). The state transitions in the SSIT are not probabilistic, and instead are contingent on (deterministically) choice actions, and the FPEQ model is a model-free algorithm that incorporates information about reward predictions from the anticipated next state using a TD-like learning rule. Therefore, we expected the cache-based model of learning reward predictions with fictive learning signals previously implemented would provide the best fit (Tobia et al., 2014). To foreshadow the results, the FPEQ model provided the best fit among the models (see Table 1) for both young and older age groups, and so details of WSLS-Q and model-based Q formalizations and results are included in Appendix A.

### MRI protocol

MR images were acquired with a 3T whole-body MR system (Magnetom TIM Trio, Siemens Healthcare) using a 32-channel receive-only head coil. Structural MRI were recorded from each participant using a T1 weighted MPRAGE sequence (voxel

resolution of  $1 \times 1 \times 1 \text{ mm}^3$ , coronal orientation, phase-encoding in left-right direction, FoV =  $192 \times 256 \text{ mm}$ , 240 slices, 1,100 ms inversion time, TE = 2.98 ms, TR = 2,300 ms, and  $90^\circ$  flip angle). Functional MR time series were recorded using a T2\* GRAPPA EPI sequence with TR = 2380 ms, TE = 25 ms, anterior-posterior phase encode, 40 slices acquired in descending (noninterleaved) axial plane with  $2 \times 2 \times 2 \text{ mm}^3$  voxels ( $204 \times 204 \text{ mm}$  FoV; skip factor = .5).

### MRI data processing

Structural and functional MR image analyses were conducted in SPM8 (Wellcome Department of Cognitive Neurology, London, UK). Anatomical images were segmented and transformed to Montreal Neurological Institute (MNI) standard space using DARTEL. Functional images were corrected for slice-timing acquisition offsets, realigned and corrected for the interaction of motion and distortion using `unwarp` toolbox, coregistered to anatomical images and transformed to MNI space using DARTEL, and smoothed (8 mm FWHM).

Event-related first-level analyses included regressors for each stimulus event shown in Fig. 1, and a set of parametric modulators for trial-specific variables and computational model parameters. Specifically, the first level GLM included a regressor marking the onset of the choice/anticipation event, outcome event, counterfactual outcome event, the state transition event, and one regressor marking the participant's response. The choice value (0€–3€) of the investment was used as a parametric modulator (modeled during the choice/anticipation event), and FPEQ model computed parametric modulators included the time series of Qsa for the selected action (modeled at the choice/anticipation event), TD prediction error (modeled at the outcome event), and FPE (modeled at the counterfactual outcome event). Thus, the TD prediction error and FPE were modeled as separate events by design, and the FPE was further divided into two sets of trials for  $f^+$  and  $f^-$  (all trials are accounted for in the model). On each trial, multiple FPEs are produced because there are multiple forgone alternatives. The FPE associated with the most optimal forgone action was used as the parametric modulator in this analysis. Coincident parametric modulators were serially orthogonalized as by default in SPM8 (i.e., the Qsa was orthogonalized with respect to the choice value).

The second level analyses contrasted the group average parameter estimates of the first level regressors in separate one-way ANOVAs (young control vs. elderly participants). All results are reported at a whole-brain cluster threshold correction using AlphaSim (Forman et al., 1995), which was performed with 2-mm isotropic voxels in MNI space, a smoothness of 8-mm, and a voxel threshold of  $p < .001$ , for 1,000 iterations, and reported that a cluster of 125 voxels achieved a statistical correction to  $p < .05$ . The results sections reports *Z*-transformed *t* values (from second level parameter

**Table 1** Model fitting parameters

| Model  | BIC<br>Mean (SD) |                | -LL<br>Mean (SD) |                |
|--------|------------------|----------------|------------------|----------------|
|        | Young            | Elderly        | Young            | Elderly        |
| StdQ   | 521.12 (120.57)  | 654.45 (15.96) | 252.34 (60.28)   | 319.01 (7.98)  |
| FPEQ   | 488.35 (107.65)  | 599.71 (65.08) | 230.47 (53.82)   | 286.16 (32.54) |
| ModelQ | 516.78 (138.33)  | 652.56 (18.87) | 247.43 (69.17)   | 315.32 (9.46)  |
| WSLS Q | 502.55 (113.99)  | 619.19 (56.07) | 223.87 (57)      | 282.19 (28.04) |

contrasts) for the peak voxel, and its MNI coordinates, from each significant cluster.

## Results

### SSIT performance: earnings and choices

Statistical comparisons revealed significant group differences for various measures of SSIT task performance. Elderly participants earned an average of 160.21€ (+/- 101.16) and the younger age group earned 355.95€ (+/- 248.35), which was significantly different,  $t(47) = -3.2584$ ,  $p < .001$ . Over the entire task (240 trials), the elderly group wagered an average of 1.42€ (+/- 1.02) on each trial, and this was less than the younger group (2.17€, +/- .98), although it did not achieve conventional statistical significance,  $t(47) = -1.9904$ ,  $p = .054$ . Early in the task, the elderly group wagered significantly more on average (1.62 +/- .39) in the first 10 rounds (30 trials) than the young (1.26 +/- .51) adult group,  $t(44.69) = 2.7153$ ,  $p = .008$ , suggesting that young adults learned to make more HR investments whereas elderly adults learned to make more LR investments. This is consistent with prior research demonstrating that the elderly are not necessarily characterized by preferences for either high or low risk (Lim & Yu, 2015; Mata, Josef, Samanez-Larkin, & Hertwig, 2011).

### SSIT performance: task knowledge, frequency of paths and state visits

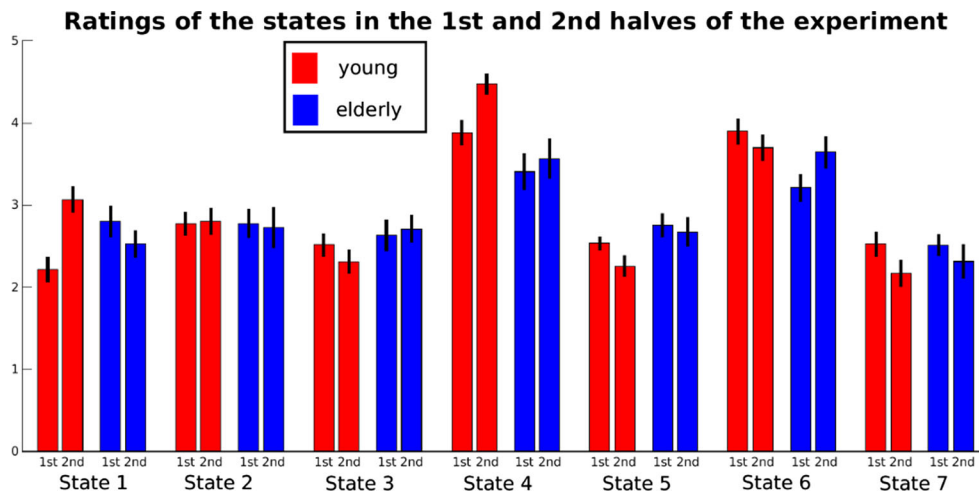
Over the entire task (80 rounds), seniors selected the path that ended in the least lucrative State 7 (53%) more frequently than the other three paths (11%, 26%, and 21% for paths ending in States 4, 5, and 6, respectively). This shows that the elderly did sometimes choose the lucrative path. However, during the final round, when participants should have sufficient knowledge of the task to perform near optimally, none of the participants in the elderly group developed a preference for the lucrative path that ends in state 4 during the final round of choices, whereas 19/30 younger participants developed a preference for the lucrative path in the final round.

Figure 3 shows the group average state ratings split into either early (1–120; i.e. first half of the experiment) or late (121–240; i.e. second half of the experiment) trials. To analyze differences in ratings for the seven states between groups and over time, an ANOVA with the between subject factor group and the within subject factors state and time (first vs. second half of the experiment) was conducted. A Greenhouse–Geisser correction to adjust the degrees of freedom was applied if necessary. The effect of group was not significant,  $F(1, 47) = .57$ ,  $p = .46$ , but the effect of states reached significance,  $F(4.64, 218.06) = 39.13$ ,  $p < .0001$ , indicating that the ratings for state 4 and 6 differed from the rating for States 1, 2, 3, 5, and 7 (Tukey HSD  $p < .000005$ ). Also, the interaction reached significance,  $F(4.4, 218.06) = 3.78$ ,  $p = .005$  indicating that the ratings for state 4 differed between groups (Tukey HSD  $p < .05$ ). This shows that the young group rated the most lucrative state (4) more highly than the elderly group, but that the elderly group rated States 4 and 6 (the two most lucrative states) more highly than the other states, showing that they had acquired some declarative knowledge about the task. The effect of time was not significant,  $F(1, 47) = .59$ ,  $p = .45$ , but the state and time interaction was significant,  $F(5.17, 243.22) = 5.30$ ,  $p < .00005$ , indicating a steeper increase for States 1 and 4 (Tukey HSD  $p < .005$ ). Also the interaction of group, state and time reached significance,  $F(5.17, 243.22) = 7.75$ ,  $p < .0001$ . To specify this interaction, post hoc group  $\times$  time ANOVAs for each state were computed and reached significance for State 1,  $F(1, 47) = 30.24$ ,  $p < .00001$ , and State 6,  $F(1, 47) = 10.21$ ,  $p < .005$ , but reached only a trend toward significance for State 4,  $F(1, 47) = 2.92$ ,  $p < .1$ . This indicates that ratings of States 1 and 4 increased over time more in the young than in the elderly, whereas it was the opposite for State 6. Critically, the ratings of the elderly also increased over time for the lucrative States 4 (although slower than in the young) and 6 showing that they acquired declarative knowledge about the states.

### Computational modeling: model fit comparisons

Table 1 presents the mean BIC scores and –LL values relating the goodness of each model's fit for the young and elderly age





**Fig. 3** State-specific ratings of experienced win frequencies for the young and elderly adult age groups are shown for the first and second halves of the SSIT. Data for the young group are shown in red, and data for the elderly group are shown in blue. For each state, the left bar shows the state rating averaged for the first half of trials, and the right bar shows

the state rating averaged for the second half. It can be seen that States 1 and 4 increased in rating for the young group but that States 3, 4, and 6 increased in rating for the elderly group, whereas the increase in ratings for State 4 was less steep. For both groups, States 4 and 6 were rated most highly. Error bars are standard error of the mean. (Color figure online)

groups. BIC scores penalize the model fit for the number of parameters to avoid over fitting. Lower BIC scores indicate better model fit. The FPEQ model had the lowest BIC scores of all models tested including the model-based Q and WLS-Q models. The BIC scores for the different models were compared with an ANOVA with model as a within-subject factor (StdQ vs. FPEQ vs. Model based vs. WLS) and group as a between-subject factor. The effects of group,  $F(1, 47) = 18.78$ ,  $p < .0001$ , and model reached significance,  $F(1.22, 57.24) = 15.93$ ,  $p < .000001$ , indicating lower BICs for the FPEQ than all other models (Tukey HSD all  $ps < .05$ ). The interaction of model and group did not reach significance,  $F(1.22, 57.24) = 1.14$ ,  $p = .34$ . However, the model fits shown in Table 1 are worse for the elderly group, and they are also less variable, for all models. This may be due to the fact that there are fewer participants in the elderly group, and they behaved more uniformly with regard to their pattern of choice in the SSIT. Together these could have led to a worse model fit that was less variable within the elderly group. We directly compared the BIC scores for the FPEQ and the nested StdQ to determine whether FPEs significantly contributed. The FPEQ model BIC score is significantly smaller than the StdQ model for the elderly subjects ( $t = 3.28$ ,  $df = 18$ ,  $p = .0021$ , paired  $t$  test) as well as for the young subjects ( $t = 2.62$ ,  $df = 29$ ,  $p = .0069$ ). Since the FPEQ model nests the StdQ model, we further compared their goodness of fit with the likelihood ratio test. For both young and elderly subjects, the FPEQ model fit the respective behavioral data significantly better than StdQ model. The likelihood ratio test statistic and  $p$  value averaged across young subjects are:  $\chi^2 = 43.7$ ,  $p = 3e^{-10}$ , and for the elderly subjects they are:  $\chi^2 = 65.7$ ,  $p = 5e^{-15}$ . Since the FPEQ model nested the StdQ model, the improvements in model fit obtained from three independent tests that penalize for

additional free parameters can be attributed directly to the inclusion of fictive error signals.

### Computational modeling: age differences for FPEQ model parameters

Table 2 presents a comparison of the FPEQ models' best fitting parameters for the younger and elderly groups. Age group differences were assessed with a two-sample  $t$  test ( $df = 47$ , two-tailed). Both age groups similarly weighted TD prediction errors (TD Learn Rate) and counterfactual losses (FPE+ Learn Rate). There was a significant difference between younger and elderly groups for the learning rate associated with counterfactual gains (FPE- Learning Rate) indicating that counterfactual gains had a stronger influence on action-specific expected values for elderly than for young adult participants. In addition, there was a significant difference between age groups for the fitted inverse temperature parameter ( $\beta$ ), indicating that younger adults made choices that exploited their task-specific acquired knowledge more so than elderly adults, and therefore the model was more stable. The age group average model computed estimates of trial-by-trial expected value for the selected action (Qsa) and absolute reward prediction errors (Abs. TD Error) were not significantly different.

Table 3 shows the correlations between the FPEQ model parameters and total earnings for each age group. All three of the learning rate parameters were significantly correlated with total earnings for the young group, but there were no significant correlations for the elderly group. Importantly, the correlation between FPE+ (counterfactual gains) and total earnings was negative, and the correlation with FPE- (counterfactual losses) was positive, showing that stronger weighting of counterfactual gains was associated with poorer SSIT performance.

**Table 2** FPEQ model parameters for elderly and young age groups

| FPEQ Parameter  | Elderly Mean (SD) | Young Mean (SD) | Elderly > Young <i>t</i> -val | Significance <i>p</i> val |
|-----------------|-------------------|-----------------|-------------------------------|---------------------------|
| Decay rate      | .675 (.40)        | .813 (.33)      | -1.26                         | .21                       |
| Exploitation    | 6.00 (3.7)        | 9.39 (4.6)      | -2.82                         | .007**                    |
| TD Learn Rate   | .18 (.20)         | .12 (.08)       | 1.19                          | .25                       |
| FPE+ Learn Rate | .028 (.06)        | .029 (.03)      | -.05                          | .95                       |
| FPE- Learn Rate | .37 (.36)         | .16 (.23)       | 2.22                          | .03*                      |
| QSA             | .33 (.03)         | .35 (.02)       | -.49                          | .63                       |
| Abs. TD Error   | .37 (.06)         | .41 (.07)       | -1.76                         | .08                       |

The sign of the correlations in the elderly group matched the signs of those in the young group, but none were statistically significant, which may symptomatic of the weaker model fit and stability parameter for the elderly group reported above.

In addition, in order to examine whether the group difference for the FPEs were related to task performance, we simulated the FPEQ model using the best fitting parameters from the young and elderly groups, respectively. Appendix B (see Figure S1) shows the young group increasingly chose the most lucrative Path 1 (blue line), whereas the elderly group increasingly chose the least lucrative Path 4 (purple line), and that model simulations yoked to the young and elderly groups' best fitting parameters reproduced their respective preferences for Paths 1 and 4. This shows that overweighting FPE- (counterfactual gains) is indeed disadvantageous for decision making in a simulated SSIT.

### Neuroimaging results

Figure 4 shows the results of the model-based fMRI analyses for the FPEQ model quantities expected value (Qsa), prediction error (TD), and fictive prediction errors (FPE+ and FPE-).

#### FPEQ: expected value (Qsa)

Figure 4 (top left) shows a significantly different correlation between Qsa for young and elderly in the vmPFC ([14 42 -16],  $z = 3.19$ ). The younger adult group showed a significant positive correlation between Qsa and neural activity in the

vmPFC, but the older adult group showed a negative correlation that was not statistically significant. This shows that reward predictions were coupled to fronto-striatal activity in young adults, but not in older adults.

#### FPEQ: fictive prediction errors (FPE+ & FPE-)

The contrast between age groups for the FPE+ parametric modulator did not reveal significant differences between groups; however, there was significant neural activity correlated negatively with FPE+. We used the minimum statistic conjunction null (Nichols, Brett, Andersson, Wager, & Poline, 2005) analysis to determine if the response to FPE+ was significantly negative for both age groups in an overlapping area of the striatum. The conjunction tests whether both groups show a similar significant effect in an overlapping region. Indeed, both age groups demonstrated a negative correlation with FPE+ (counterfactual loss) in the ventral striatum ([16 4 -6],  $z = 2.89$ ) as indicated by the significant conjunction analysis shown in Fig. 4 (top middle). While it can be seen that the elderly group again showed a stronger effect than the younger group, it was not significantly different. The contrast examining the effect of age on FPE- (counterfactual gain) failed to retain any clusters with the whole-brain cluster threshold statistical correction. Because prior research has shown that the ventral striatum is involved in processing counterfactual outcome signals (Büchel et al., 2011; Lohrenz et al., 2007) we tested for a significant effect with a small volume correction, which is less conservative than the whole-brain correction. We used a voxelwise small volume correction by placing a search sphere with an 8 voxel (16 mm) radius around the right and left nucleus accumbens (MNI: [12 8 8; -12 8 -8]). The elderly group showed a stronger (negative) modulation in the left ventral striatum (see Fig. 4, top right) with a peak difference at ([-14 8 -12],  $z = 3.03$ ;  $p_{FWE} = .023$ ;  $p = .001$ , uncorrected) in the left anterior putamen.

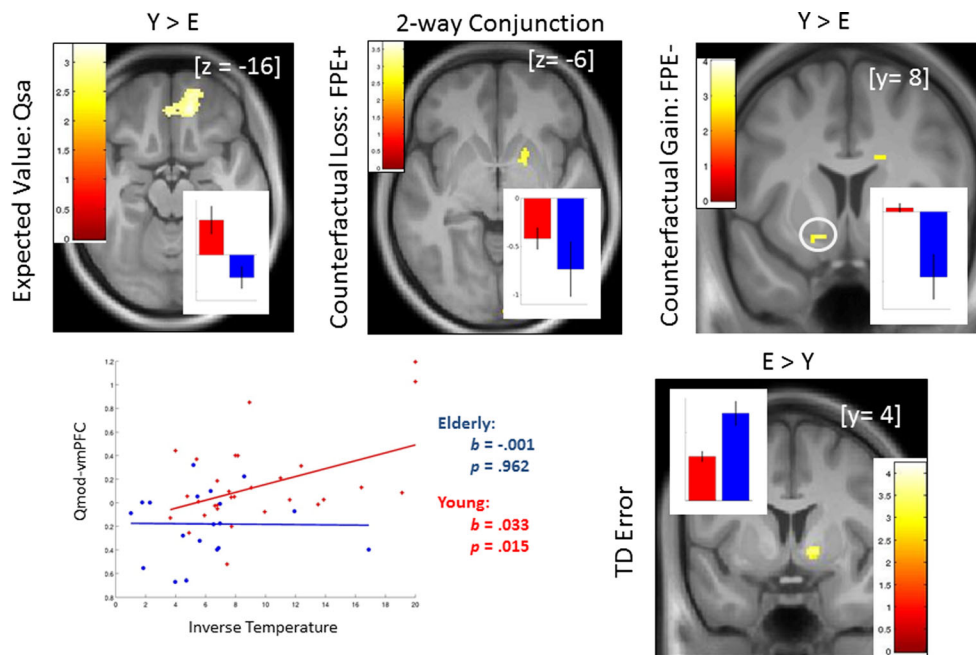
#### FPEQ: reward prediction error (TD)

Both age groups demonstrated significant positive correlations between TD reward prediction errors and neural activity

**Table 3** Correlation of model parameters and total earnings

|                 | Elderly <i>r</i> ( <i>p</i> ) | Young <i>r</i> ( <i>p</i> ) |
|-----------------|-------------------------------|-----------------------------|
| Decay Rate      | .024 (.92)                    | .11 (.56)                   |
| TD Learn Rate   | -.01 (.96)                    | -.42 (.02)*                 |
| FPE+ Learn Rate | -.36 (.13)                    | -.64 (.001)**               |
| FPE- Learn Rate | .23 (.34)                     | .39 (.03)*                  |
| Exploitation    | .10 (.68)                     | .33 (.08)                   |

\* $p < .05$ ; \*\* $p < .001$



**Fig. 4** Age differences in BOLD signal modulation by expected value of the chosen action (Qsa parameter) on each trial (left), counterfactual gain (FPE-; top right) and TD reward prediction error (bottom right). Both age groups showed a similar negative modulation in the ventral striatum for counterfactual losses (FPE+) indicated by a significant conjunction analysis (top center). Inset bar graphs depict the direction and amplitude of the effect on BOLD signal, with data from the young group shown in the red bars and data from the elderly group in the blue bars. Data in the graphs are the group's mean beta value ( $\pm$  standard error of the mean). Images are shown on the MNI template at whole brain cluster correction

to  $p < .05$ . The group contrast for FPE- was significant with a small volume correction as described in the results, and is shown at  $p < .005$  uncorrected for visualization. *Note.* Y = young; E = elderly. The scatter graph at the bottom left illustrates the relationship between the FPEQ model inverse temperature parameter ( $x$ -axis) and parametric modulation in the vmPFC by expected value ( $y$ -axis). Data and best fit regression line for the elderly are shown in blue. Data and best fit regression line for the young adults are shown in red. Lines show a positive relationship for young adults and no relationship for elderly adults. (Color figure online)

in the ventral striatum. However, the elderly group showed a stronger correlation ( $[14\ 4\ -4]$ ,  $z = 3.49$ ) than the younger group in the right ventral striatum (see Fig. 4, bottom right), suggesting that elderly participants' neural activity is more strongly coupled to reward prediction errors.

#### FPEQ: Exploitation, model stability, and expected value

Additionally, we examined the relationship between the FPEQ model inverse temperature parameter,  $\beta$ , and group differences in neural activity involved in valuation as identified by the model (Qsa/vmPFC, TD error/ventral striatum, FPE-/ventral striatum). We conducted separate linear regression analyses for the young and elderly groups' data using the MATLAB function *robustfit* with bisquare weighting to suppress the influence of outliers on the linear fit. For each analysis, the FPEQ model  $\beta$  was the independent variable predicting modulated neural activity (first level GLM parameters for Qsa, TD error, or FPE-, in their respective peak voxels reported above) as the dependent variable. The  $\beta$  parameter significantly predicted Qsa-correlated neural activity in the vmPFC for the young adult group,  $b = .033$ ,  $t(28) = 2.58$ ,  $p = .015$ , but not for the elderly group,  $b = -.001$ ,  $t(17) = -.048$ ,  $p = .962$ . These results are illustrated in the scatter and line plot at bottom left

of Fig. 4. The results for factual and counterfactual prediction errors were nonsignificant, all  $ps > .05$ .

#### Discussion

Learning from fictive prediction errors is an important part of reward prediction, and the neural system for reward prediction from standard reward and fictive prediction errors involves a fronto-striatal circuit (Chiu et al., 2008; Lohrenz et al., 2007; Tobia et al., 2014). This study investigated whether age affects decision making based on reward predictions learned with fictive prediction errors (FPEs). We found that older participants performed poorly on the SSIT compared to younger adults, weighted FPEs differently than younger adults, and showed different correlations between neural activity in the fronto-striatal circuit and reward predictions, reward prediction errors and FPEs. The FPEQ model explained the data from both age groups better than three other models previously used in the literature, which shows that decision making guided by reward predictions, rather than outcome-response strategies, such as win-stay-lose-shift, was the best fit for the elderly. The reward predictions, prediction errors and FPEs from the FPEQ model, were correlated with neural activity

in the fronto-striatal circuit previously implicated in age-related differences in processing reward predictions and reward prediction errors (Chowdhury et al., 2013; Eppinger et al., 2012; Rademacher et al., 2014; Samanez-Larkin et al., 2007, 2014; Vink et al., 2015). Importantly, age did not significantly affect TD learning rate nor the average experienced TD error. Taken together, these results support the interpretation that poor SSIT performance by the elderly group was related to altered behavioral and neural responsiveness to counterfactual outcomes.

### **Disadvantageous overweighting of counterfactual gains in the elderly**

Evidence from the FPEQ model and fMRI data agree that the elderly group was more responsive, behaviorally and neurally, to counterfactual gains – which were not associated with advantageous future-oriented decision making by design in the SSIT. The elderly group was more sensitive to FPE- than the younger group according to the significantly increased FPE- learning rate parameter, and the fMRI results showed a stronger neural response (negative correlation) in the left posterior striatum to the FPE- event. In addition, the elderly adults demonstrated a similar FPE+ response in the right ventral striatum as the young, and the FPEQ model weighted FPE+ similarly for both groups, showing no differences in behavioral or neural responsiveness to counterfactual losses between groups.

Previous research has demonstrated that healthy elderly adults are less affected by counterfactual outcomes (Brassen et al., 2012) and actually utilize decision making strategies that don't rely on reward predictions, such as win-stay-lose-shift (Worthy & Maddox, 2012). The results of the model fitting shows that elderly adults SSIT performance was more likely to be guided by reward predictions than a strategy that doesn't utilize reward predictions. Furthermore, the FPEQ model results indicate that counterfactual gains and losses significantly impacted reward predictions and decision making in both age groups. Three independent tests that penalize a model for free parameters verified that the FPEQ model fit the data better than the StdQ model (as well as a model-based Q and win-stay-lose-shift Q model; see Table 1 for model fits and Appendix A for model details). The StdQ model nested the FPEQ model such that the models are equivalent if the fictive prediction errors fail to make a contribution (learning rate parameters are 0). As such, any improvement in model fit can be attributed to incorporating fictive prediction errors, and the significant improvement identified by three separate tests emphasizes the importance of counterfactual information for generating reward predictions in accounting for SSIT performance.

Our results also show that elderly adults are significantly more sensitive to counterfactual gains than young adults,

which may account for performance differences between the age groups. The elderly group weighted counterfactual gains significantly more than the young adults, but there was no group difference for counterfactual losses. The larger learning rate for counterfactual gains by the elderly means that the expected values of disadvantageous LR investments are disproportionately increased compared to the young adults. At the same time that counterfactual gains increase values for LR investments (i.e., any losing trial with an investment greater than 0), the expected values of advantageous HR investments are unadjusted, thereby leading to a greater likelihood of overvaluing and choosing a nonlucrative path through the SSIT, as demonstrated by the elderly group in this study.

This finding seems in conflict with prior research showing healthy older adults are insensitive to counterfactual outcomes (Brassen et al., 2012). However, this difference is likely related to substantive differences in the paradigms between the current and previous research. In the current research, counterfactual outcomes are informative with respect to subsequent choices in that they clue participants to the true value of the decision path (i.e., choice-dependent task) allowing acquisition of reward predictions for actions not selected, whereas previous research employed a task with discrete trials whose outcome was irrelevant (i.e., choice-independent task) for future choice (Brassen et al., 2012). Thus, our findings are in agreement with Worthy and Maddox (2014) concerning choice-dependent and choice-independent tasks; when counterfactual prediction errors signals can be used for learning reward predictions to make choice-dependent decisions, elderly adults are indeed responsive. Furthermore, the incongruent findings between our results and Brassen and colleagues' findings seem to indicate that elderly participants are insensitive in one context (choice-independent), and overly sensitive in another context (choice-dependent), to counterfactual gains and losses, and that in either case it is significantly different from the behavior of younger adults. This suggests that the aging system does not adapt optimally to either task context, and instead settles at extremes rather than an optimal intermediate state.

### **Counterfactual gains and reduced choice consistency**

The results reported above suggest that age-related differential weighting and neural processing of counterfactual gains, but not counterfactual losses, may be related to relatively poor SSIT performance by an effect on the stability of reward predictions. The group analysis of the FPEQ model revealed only two parameters that were significantly different between younger and elderly adults: the FPE- learning rate for counterfactual gains discussed above, and the inverse temperature parameter,  $\beta$ , indicating the influence of acquired knowledge on behavioral stability. When the value of  $\beta$  is high, more knowledge was exploited during choice and choice patterns



become more stable, and low values are related to exploration and prolonged learning. A more stable pattern of choices in the SSIT can lead to greater success (i.e., accrued more reward) if it follows a lucrative path, or less success if it follows a nonlucrative path. The significantly reduced  $\beta$  values in the elderly group are suggestive that more choices were dedicated to exploration of the task, leading to reduced behavioral stability and a prolonged learning period. This suggests that elderly participants' performance should improve given sufficient trials. The SSIT presents a large number of trials, as well as interim probes of task-related declarative knowledge (excluding knowledge of state transition rules). Results of the ANOVAs indicated that the elderly group demonstrated sufficient declarative knowledge of the task (that States 4 and 6 on average were rated most highly); however, the interaction from the three-way ANOVA showed that the young and elderly groups changed their values for States 1, 4, and 6 differently. The ratings for States 1 and 4 increased for the young group, suggesting they had knowledge that these states led to a higher expected value, and is consistent with their preference for the most lucrative path during the final round of trials. The elderly group showed, instead, an increase in rating for State 6, a slower increase in rating for State 4, and no increase in rating for State 1. The increase in rating in State 1 is important because a HR wager in this state is required to advance along the most lucrative path, and, as such, none of the elderly preferred the most lucrative path during the final 10 rounds of the task (Figure S1 in Appendix B also shows that the elderly group preferred the least lucrative path throughout the experiment). The group difference in ratings and the lack of improvement in performance by the late rounds of the task in the elderly group suggests that counterfactual outcomes may be involved in generating choice inconsistency and reduced model stability in the elderly.

Over-weighting counterfactual gains may reduce choice consistency by altering the coupling of reward predictions and neural activity in the vmPFC because of conflict between knowledge that a path can produce a large payout versus the expected value of an action on a particular trial. Highly valued LR investments, such as those influenced by counterfactual gains, are in conflict with acquired knowledge that states or paths produce larger payouts because they preclude the advantageous state transitions necessary to follow the lucrative paths. This could lead to less stable choice patterns and a reduced  $\beta$  parameter if resolution of such choice conflict vacillates between exploiting knowledge for a given pattern of choices by placing HR investments (i.e., a path) on some trials, and exploiting knowledge that is strongly influenced by counterfactual gains (i.e., expected value) by placing LR investments on other trials. As such, elderly participants may not have been able to consistently overcome the inflated expected values for LR investments stemming from their over-weighting of counterfactual gains. Furthermore, increasing

values of  $\beta$  predicted a stronger relationship between Qsa and neural activity in the vmPFC for the young adult group only. The lack of a relationship in the elderly group may be due to vacillation between choosing actions guided by FPEQ reward predictions (LR wagers for the elderly group) versus actions that are not highly valued by the model (i.e., HR wagers). However, it is necessary to take this interpretation with caution because the softmax beta captures anything that is not related to reward maximization (including model misfit and noise) and may therefore have multiple (possibly poorly defined) interpretations.

### Reward prediction errors in the ventral striatum

The TD error was more strongly correlated with neural activity in the ventral striatum than the young adults, but there were no differences between age groups for the experienced TD error or TD learning rate. This is consistent with prior work showing altered neural processing of reward prediction errors in the elderly (Eppinger, Schuck et al., 2013), but deviates in that elderly adults generally show a combination of reduced behavioral and neural responsiveness to reward prediction errors. Moreover, it reflects a general trend in our data by which older adult's neural activity is more strongly correlated to consequences, factual and counterfactual, than young adults. Such subtle differences in neural processing, in the absence of measurable behavioral differences, may be indicative of age-related changes in neural activity at a smaller scale, such as age-altered dopamine or serotonin functioning (Backman et al., 2000; Chowdhury et al., 2013; Goldberg et al., 2004; Lamar et al., 2009; Matuskey et al., 2012; Volkow et al., 1998). In our previous research, we found that altered dopamine levels were related to different correlations between TD error and neural activity in the ventral striatum but there was no effect on learning from TD reward prediction errors (Tobia et al., 2014). In addition, we also found that altered serotonin levels were associated with a reversed correlation between reward predictions (Qsa) and neural activity in the vmPFC, but there was no effect on reward predictions per se (Tobia et al., 2014). Together, these suggest that the model-based experimental design may be sensitive enough to identify age-related effects of neurotransmitter-specific activity on the component processes of reward prediction and prediction error even when there are no detectable behavioral differences (Tobia et al., 2014). The age effect on correlations between neural activity and reward prediction errors (TD error) and reward predictions (Qsa) implicates age-altered functioning of dopamine and serotonin systems in the ventral striatum and the vmPFC, respectively. Furthermore, this tendency to overactivate for prediction errors relative to the young adults may reflect a compensatory response in the elderly due to age-altered efficiency in neural processing circuits. This may explain why there were age

differences in neural activity, despite the lack of detectable behavioral age differences in reward prediction and responsiveness to reward prediction errors.

### Limitations and future research

There are several limitations of this study that require taking our data and conclusions cautiously. First, the model fits were not equivalent between age groups, suggesting that the FPEQ model, although the best fitting among the models we implemented, may not provide a good comparison between the age groups. However, differences in model fit are not unusual and difficult to avoid when comparing groups. Future research could examine a larger sample of older adults and select a subset of those with good model fits for subsequent analyses. This speaks to a larger problem, where the elderly may represent a heterogeneous population for which no single model is the best predictor of behavior, thus requiring multiple models to be integrated in order to fully characterize age effects. Additionally, our results could be related to competition between cache-based and model-based systems (Daw et al., 2011) in the elderly. However, a model-based Q model was less good fit than the FPEQ model, consistent with our expectation that the deterministic state transitions in the SSIT (a choice-dependent task) would be sufficiently explained by action-specific reward predictions, rather than decision making guided by state-transition predictions per se. Finally, the mediating effect of altered working memory in the elderly could impair task performance. Other research shows that age-related effects of working memory on decision making in multialternative choice tasks do not fully explain task performance differences between age groups (Eppinger, Walter, et al., 2013). In addition, our participants were thoroughly screened for intellectual deficits, and so it is unlikely to greatly influence the results of this study.

### Summary

The results of this study demonstrate that age significantly impacts decision making based on reward predictions learned with FPEs. Elderly adults exhibited greater behavioral and neural responsiveness to counterfactual gains (FPE-) than younger adults. Task-related activity in the fronto-striatal circuit was also affected by age, where older adults generally showed stronger responses to factual and counterfactual consequences in the striatum, but a non-significant correlation with reward predictions in the vmPFC. Our findings are in general agreement with previous literature showing that relatively poor decision making in multialternative choice tasks in older adults is related to aberrant reward prediction and prediction error processing in this fronto-striatal circuit (Chowdhury et al, 2013; Eppinger, Walter et al., 2013; Eppinger, Nystrom, et al., 2012; Eppinger, Schuck, et al.,

2013; Grady, 2012; Hedden & Gabrieli, 2004; Rademacher, Saalma, Grunder, & Sprecklemeyer, 2014; Samanez-Larkin et al., 2007; Samanez-Larkin et al., 2011; Samanez-Larkin et al., 2014; Samanez-Larkin & Knutson, 2015; Vink et al., 2015; ), and extends these findings to include altered processing of counterfactual consequences. Finally, this study advocates the model-based fMRI experimental design to study both behavioral and neural systems for decision making between age groups but cautions that model fits be carefully taken into account when drawing conclusions about group differences. Future research should design tasks that allow application of multiple models, or integrated models, to account for the diversity of decision making mechanisms and could consider how counterfactual gains and losses affect arbitration between model-based and cache-based decision-making mechanisms in aging.

**Acknowledgments** This work was funded by a Bernstein Prize for Computational Neuroscience BMBF 01GQ1006 to J. G., as well as BMBF 01GQ0912, BMBF 01GQ0911, and DFG GRK 1589/1

### References

- Backman, L., Ginovart, N., Dixon, R. A., Wahlin, T. B., Wahlin, A., Halldin, C., & Farde, L. (2000). Age-related cognitive deficits mediated by changes in the striatal dopamine system. *American Journal of Psychiatry*, *157*(4), 635–637.
- Barbey, A. K., Krueger, F., & Grafman, J. (2009). Structured event complexes in the medial prefrontal cortex support counterfactual representations for future planning. *Philosophical Transactions of the Royal Society B*, *364*, 1291–1300.
- Bennet, I. J., Madden, D. J., Vaidya, C. J., Howard, D. V., & Howard, J. H., Jr. (2010). Age-related differences in multiple measures of white matter integrity. *Human Brain Mapping*, *31*(3), 378–390.
- Borst, J. P., & Anderson, J. R. (2013). Using model-based functional MRI to locate working memory updates and declarative memory retrievals in the fronto-parietal network. *Proceedings of the National Academy of Science*, *110*, 1628–1633.
- Brassen, S., Gamer, M., Peters, J., Gluth, S., & Büchel, C. (2012). Don't look back in anger! Responsiveness to missed chances in successful and unsuccessful aging. *Science*, *336*, 612–614.
- Brickman, A. M., Habeck, C., Zarahn, E., Flynn, J., & Stern, Y. (2006). Structural MRI covariance patterns associated with normal aging and neuropsychological functioning. *Neurobiology of Aging*. doi: 10.1016/j.neurobiolaging.2005.12.016
- Büchel, C., Brassen, S., Yacubian, J., Kalisch, R., & Sommer, T. (2011). Ventral striatal signal changes represent missed opportunities and predict future choice. *NeuroImage*, *57*, 1124–1130.
- Burzynska, A. Z., Preuschof, C., Backman, L., Nyberg, L., Li, S.-C., Lindenberger, U., & Heekeren, H. R. (2010). Age-related differences in white matter microstructure. *NeuroImage*, *49*, 2104–2112.
- Byrne, R. M. J. (2002). Mental models and counterfactual thoughts about what might have been. *TICS*, *6*, 426–430.
- Chandler, M. J., Lacritz, L. H., Hyman, L. S., Barnard, H. D., Allen, G., Deschner, M., ... Cullum, C. M. (2005). A total score for the CELRD neuropsychological battery. *Neurology*, *65*, 102–106.

- Chiu, P. H., Lohrenz, T. M., & Montague, P. R. (2008). Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. *Nature Neuroscience*, *11*, 515–520.
- Chowdhury, R., Guitart-Masip, C. L., Dayan, P., Huys, Q., Düzel, E., & Dolan, R. J. (2013). Dopamine restores reward prediction errors in old age. *Nature Neuroscience*, *16*, 648–653.
- Coricelli, G., & Rustichini, A. (2010). Counterfactual thinking and emotions: Regret and envy learning. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *365*, 241–247.
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*, 1204–1215.
- De Brigard, F., Addis, D. R., Ford, J. H., Schacter, D. L., & Giovanello, K. S. (2013). Remembering what could have happened. *Neuropsychologia*, *51*, 2401–2414.
- Doll, B. B., Simon, D., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology*, *22*, 1075–1081.
- Eppinger, B., Heekeren, H. R. & Li, S. C. (2015). Age-related prefrontal impairments implicate deficient prediction of future reward in older adults. *Neurobiology of Aging*, *36*, 2380–2390.
- Eppinger, B., Nystrom, L. E., & Cohen, J. D. (2012). Reduced sensitivity to immediate reward during decision-making in older than younger adults. *PLoS ONE*, *7*, e36953.
- Eppinger, B., Schuck, N. W., Nystrom, L. E., & Cohen, J. D. (2013). Reduced striatal responses to reward prediction errors in older compared with younger adults. *The Journal of Neuroscience*, *33*, 9905–9912.
- Eppinger, B., Walter, M., Heekeren, H. R., & Li, S.-C. (2013). Of goals and habits: Age-related and individual differences in goal-directed decision-making. *Frontiers in Neuroscience*, *7*, 253. doi:10.3389/fnins.2013.00253
- Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic Resonance in Medicine*, *33*, 636–647.
- Gläscher, J., Daw, N. D., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*, 585–595.
- Gläscher, J., Hampton, A. N., & O'Doherty, J. P. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cerebral Cortex*, *19*, 483–495.
- Gläscher, J., & O'Doherty, J. P. (2010). Model-based approaches to neuroimaging: Combining reinforcement learning theory with fMRI data. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 501–510.
- Goldberg, S., Smith, G.S., Barnes, A., Ma, Y., Kramer, E., Robeson, K., . . . Eidelberg D. (2004). Serotonin modulation of cerebral glucose metabolism in normal aging. *Neurobiology of aging*, *25*, 167–174.
- Grady, C. (2012). The cognitive neuroscience of aging. *Nature Reviews Neuroscience*, *13*, 491–505.
- Gu, X., Kirk, U., Lohrenz, T. M., & Montague, P. R. (2013). Cognitive strategies regulate fictive, but not reward prediction error signals in a sequential investment task. *Human Brain Mapping*. doi:10.1002/hbm.22433
- Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *Journal of Neuroscience*, *28*, 5623–5630.
- Hautzinger, M., Bailer, M., Worall, H., & Keller, F. (1995). *Beck-Depressions-Inventar (BDI)*. *Testhandbuch*. [Beck Depression Inventory. Test Handbook]. (2nd ed.). Bern, Switzerland: Hans Huber.
- Hedden, T. & Gabrieli, J. D. E. (2004). Insights into the ageing mind: a view from cognitive neuroscience. *Nature Reviews Neuroscience*, *5*, 87–96
- Horhota, M., Mienaltowski, A., & Blanchard-Fields, F. (2012). If only I had taken my usual route . . . : Age-related differences in counterfactual thinking. *Aging, Neuropsychology, and Cognition*, *19*, 339–361.
- Kreuzpointner, L., Lukesch, H., & Horn, W. (2013). *Leistungsprüfungssystem 2. LPS-2*. [Performance Test System 2]. Göttingen, Germany: Hogrefe.
- Lamar, M., Cutter, W. J., Rubia, K., Brammer, M., Daly, E. M., Craig, M. C., . . . Murphy, D. G. (2009). 5HT, prefrontal function and aging. *Neurobiology of Aging*, *30*, 1135–1146.
- Levens, S. M., Larsen, J. T., Bruss, J., Tranel, D., Bechara, A., & Mellers, B. A. (2014). What might have been? The role of ventromedial prefrontal cortex and lateral orbitofrontal cortex in counterfactual emotions and choice. *Neuropsychologia*, *54*, 77–86.
- Lim, K. T. K., & Yu, R. (2015). Aging and wisdom: Age-related changes in economic and social decision making. *Frontiers in Aging Neuroscience*. doi:10.3389/fnagi.2015.00120
- Lohrenz, T., McCabe, K., Camerer, C. F., & Montague, P. R. (2007). Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academies of Science of the United States of America*, *104*, 9493–9498.
- Mata, R., Josef, A. K., Samanez-Larkin, G. R., & Hertwig, R. (2011). Age-differences in risky choice: A meta-analysis. *Annals of the New York Academy of Science*, *1235*, 18–29.
- Matuskey, D., Pittman, B., Planeta-Wilson B., Walderhaug, E., Henry, S., Gallezot, J. D. & Neumeister, A. (2012). Age effects on serotonin receptor 1B as assessed by PET. *Journal of Nuclear Medicine*, *53*, 1411–1414.
- McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., Montague P. R. (2004). Neural correlates of behavioral preference for culturally familiar drinks. *Neuron*, *44*, 379–87.
- Montague, P. R., King-Cassas, B., & Cohen, J. D. (2006). Imaging valuation models of choice. *Annual Review of Neuroscience*, *29*, 417–448.
- Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, *25*, 653–660.
- O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Current Opinion in Neurobiology*, *14*, 769–776.
- Pfefferbaum, A., Adalsteinsson, E., & Sullivan, E. V. (2005). Frontal circuitry degradation marks healthy adult aging: Evidence from diffusion tensor imaging. *NeuroImage*, *26*, 891–899.
- Rademacher, L., Saalma, A., Grunder, G., & Sprecklemeyer, K. N. (2014). Differential patterns of nucleus accumbens activation during anticipating of monetary and social reward in young and old adults. *SCAN*, *9*, 825–831.
- Samanez-Larkin, G. R., Gibbs, S. E., Khanna, K., Nielsen, L., Carstensen, L. L., & Knutson, B. (2007). Anticipation of monetary gain but not loss in healthy older adults. *Nature Neuroscience*, *10*, 787–791.
- Samanez-Larkin, G. R., & Knutson, B. (2015). Decision making in the ageing brain: Changes in affective and motivational circuits. *Nature Reviews: Neuroscience*, *16*, 278–289.
- Samanez-Larkin, G.R., Levens, S.M., Perry, L.M., Dougherty, R.F. & Knutson, B. (2012). Frontostriatal white matter integrity mediates adult age differences in probabilistic reward learning. *Journal of Neuroscience*, *32*, 5333–5337.
- Samanez-Larkin, G. R., Kuhnen, C. M., Yoo, D. J., & Knutson, B. (2010). Variability in nucleus accumbens activity mediates age-

- related suboptimal financial risk taking. *The Journal of Neuroscience*, *30*, 1426–1434.
- Samanez-Larkin, G. R., Wagner, A. D., & Knutson, B. (2011). Expected value information improves financial risk taking across the adult life span. *Social Cognitive Affective Neuroscience*, *6*, 207–217.
- Samanez-Larkin, G. R., Worthy, D. A., Mata, R., McClure, S. M., & Knutson, B. (2014). Adult age differences in frontostriatal representation of prediction error but not reward outcome. *Cognitive, Affective, & Behavioral Neuroscience*, *14*, 672–682.
- Sommer, T., Peters, J., Gläscher, J., & Büchel, C. (2009). Structure-function relationships in the processing of regret in the orbito-frontal cortex. *Brain Structure and Function*, *213*, 535–551.
- Thalman, B., Monsch, A. U., Schneitter, M., Bernasconi, F., Aebi, C., Camachova-Davet, Z., & Staehelin, H. B. (2000). The CELRD neuropsychological assessment battery (CELRD-NAB): A minimal dataset as a common tool for German-speaking Europe. *Neurobiology of Aging*, *21*, 30.
- Tobia, M. J., Guo, R., Schwarze, U., Boehmer, W., Gläscher, J., Finckh, B.,... Sommer, T. (2014). Neural systems for valuation and choice with counterfactual learning signals. *NeuroImage*, *89*, 57–69.
- Vink, M., Kleerekooper, I., van den Wildenberg, W. P. M., & Kahn, H. R. (2015). Impact of aging on frontostriatal reward processing: Impact of aging. *Human Brain Mapping*, *36*, 2305–2317.
- Volkow, N., Gur, R. C., Wang, G. J., Fowler, J. S., Moberg, P. J. Ding, Y. S. Hitzemann, R.,... Logan, J. (1998). Association between decline in brain dopamine activity with age and cognitive and motor impairment in healthy individuals. *American Journal of Psychiatry*, *155*, 344–349.
- Watkins, C., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*, 279–292.
- Worthy, D. A., & Maddox, W. T. (2012). Age-based differences in strategy use in choice tasks. *Frontiers in Neuroscience*, *5*. doi:10.3389/fnins.2011.00145
- Worthy, D. A., & Maddox, W. T. (2014). A comparison model of reinforcement-learning and win-stay-lose-shift decision-making processes: A tribute to W. K. Estes. *Journal of Mathematical Psychology*, *59*, 41–49.