

Testing Theories of Investor Behavior Using Neural Data

Cary Frydman, Nicholas Barberis, Colin Camerer,
Peter Bossaerts and Antonio Rangel¹

Abstract. We show that measures of neural activity provided by functional magnetic resonance imaging (fMRI) can be used to test between theories of investor behavior that are difficult to distinguish using behavioral data alone. Subjects traded stocks in an experimental market while we measured their brain activity. Behaviorally, we find that, our average subject exhibits a strong disposition effect in his trading, even though it is suboptimal. We then use the neural data to test a specific theory of the disposition effect, the “realization utility” hypothesis, which argues that the effect arises because people derive utility directly from the act of realizing gains and losses. Consistent with this hypothesis, we find that activity in an area of the brain known to encode the value of decisions correlates with the capital gains of potential trades, that the size of these neural signals correlates across subjects with the strength of the behavioral disposition effects, and that activity in an area of the brain known to encode experienced utility exhibits a sharp upward spike in activity at precisely the moment at which a subject issues a command to sell a stock at a gain.

¹ Frydman, Camerer, Bossaerts & Rangel are at Caltech; Barberis is at the Yale School of Management; Bossaerts is also at the Swiss Finance Institute. We are grateful for comments from participants at the Society for Neuroeconomics, the Fall 2010 NBER Behavioral Finance meeting, the 2010 Miami Finance Conference, the 2011 BEAM Conference, and the 2011 WFA Conference. Financial support from the National Science Foundation (A.R., C.C., C.F.) and the Betty and Gordon Moore Foundation (A.R., C.C.) is gratefully acknowledged.

Over the past twenty years, economists have accumulated a large amount of evidence on how individual investors manage their financial portfolios over time. Some of this evidence is puzzling, in the sense that it is hard to reconcile with the simplest models of rational trading (Barberis and Thaler (2003); Campbell (2006)). Theorists have responded to this challenge by constructing new models of investor behavior. Empiricists, in turn, have started testing these newly-developed models.

Most of the empirical work that tests theories of investor behavior uses field data (Barber and Odean (2000); Barber and Odean (2001); Choi et al. (2009); Grinblatt and Keloharju (2009)). A smaller set of studies uses data from laboratory experiments. The advantage of experimental data is that it gives researchers a large degree of control over the trading and information environment, which can make it easier to tease theories apart (Plott and Sunder (1988); Camerer and Weigelt (1991); Camerer and Weigelt (1993); Weber and Camerer (1998); Bossaerts and Plott (2004); Bossaerts et al. (2007)).

In this paper, we show that another kind of data, namely measures of *neural* activity taken using functional magnetic resonance imaging (fMRI) while subjects trade in an experimental stock market, can also be very useful in testing theories of investing behavior. In particular, we show that neural data can be used to test theories designed to explain the “disposition effect,” the robust empirical fact that individual investors have a greater propensity to sell stocks trading at a gain relative to purchase price, rather than stocks trading at a loss².

The disposition effect has attracted considerable attention because it has proven challenging to explain using simple rational models of trading behavior. This impasse has motivated the development of multiple competing alternative theories, both rational and behavioral (Shefrin and Statman (1985); Odean (1998); Barberis and Xiong (2009); Kaustia (2010)). One such theory, which is the focus of this paper, is the realization utility hypothesis

² See for example, Shefrin and Statman (1985), Odean (1998), Genesove and Mayer (2001), Grinblatt and Keloharju (2001), Feng and Seasholes (2005), Frazzini (2006), Jin and Scherbina (2011).

(Shefrin and Statman (1985); Barberis and Xiong (2011)). According to this theory, in addition to deriving utility from consumption, investors also derive utility *directly* from *realizing* gains and losses on the sale of risky assets that they own. For example, if an investor realizes a gain (e.g., by buying a stock at \$20 and selling it at \$40), he receives a positive burst of utility proportional to the capital gain. In contrast, if he realizes a loss (e.g., by buying a stock at \$20 and selling it at \$10), he receives a negative burst of utility proportional to the size of the realized loss. The presence of realization utility is important because, in combination with a sufficiently high time discount rate, it leads investors to exhibit a disposition effect (Barberis and Xiong (2011)).

Testing among competing theories of phenomena like the disposition effect using field or experimental data is difficult because these theories often make similar predictions about behavior (Weber and Camerer (1998) is an exception). Furthermore, it is extremely difficult, using such data alone, to carry out direct tests of the mechanisms driving behavior (e.g., of whether or not people actually receive bursts of utility proportional to realized capital gains). On the other hand, a combination of neural measurement and careful experimental design allows for direct tests of the extent to which the computations made by the brain at the time of decision-making are consistent with the mechanisms posited by different models.

In this paper, we describe the results of an fMRI experiment designed to test the hypothesis that subjects experience realization utility while trading in an experimental stock market, and that this is associated with trading patterns consistent with the disposition effect. The experiment allows us to test several behavioral and neural predictions of the realization utility hypothesis.³

Behaviorally, we find that the average subject in our experiment exhibits a strong and significant disposition effect. This stands in sharp contrast to the prediction of a simple rational

³ In this paper, we use the word “behavioral” in two different senses. Most of the time, as in the last sentence of this paragraph, we take it to mean “pertaining to behavior”. Occasionally, we take it to mean “less than fully rational” or “psychological”. It should be clear from the context which of the two meanings is intended.

trading model in which subjects maximize the expected value of final earnings. In particular, our experimental design induces positive short-term autocorrelation in stock price changes, which implies that a risk-neutral rational trader would sell losing stocks more often than winning stocks, thereby exhibiting the opposite of the disposition effect. In contrast, the strong disposition effect displayed by our subjects is consistent with the existence of realization utility effects.

When taken literally as a model of the decision-making process, the realization utility model also makes several clear predictions about the pattern of neural activity that should be observed at different times in the experiment. We describe these predictions in detail in the main body of the paper, but summarize them briefly here.

First, the realization utility model predicts that, at the moment when a subject is making a decision as to whether to sell a stock, neural activity in areas of the brain that are known to encode the value of potential actions should be proportional to the capital gain that would be realized by the trade (i.e. to the difference between the sale price and the purchase price). This prediction follows from the fact that, for an individual who experiences realization utility, the value of selling a stock depends on the associated capital gain or loss. Brain regions that have been widely shown to correlate with the value of potential actions include the ventromedial prefrontal cortex (vmPFC) and the ventral striatum (vSt)⁴.

Second, the realization utility model predicts that, across individuals, the strength of the disposition effect should be correlated with the strength of the realization utility signal in decision value areas such as the vmPFC or the vSt. This follows from the fact that a subject who is strongly influenced by realization utility should exhibit both a strong disposition effect *and* neural activity in decision value areas that is highly responsive to the associated capital gain.

Third, the realization utility hypothesis predicts that neural activity in areas that have been associated with the encoding of experienced utility (sometimes called “instantaneous

⁴ See Hsu et al. (2005), Kable and Glimcher (2007), Knutson et al. (2007), Hare et al. (2008), Kennerley et al. (2008), Chib et al. (2009), Hare et al. (2009), Hsu et al. (2009), Kang et al. (2009), Hare et al. (2010), Levy et al. (2010), Litt et al. (2010), Kang et al. (2011).

hedonics”) should increase at the moment that a subject decides to realize a capital gain. Previous research in behavioral neuroscience has shown that activity in regions of the vmPFC and the vSt also correlates with the reported level of instantaneous experienced utility⁵. This prediction is particularly interesting because it provides the most direct test of the realization utility hypothesis, and thus best illustrates the value of neural data for testing theories of financial decision-making.

Our fMRI measurements reveal patterns of neural activity that are consistent with the three neural predictions. This provides novel and strong support for the mechanisms at work in the realization utility model, and to our knowledge, provides the first example of how neural evidence can be used to test economic models of financial decision-making. We emphasize that the results do not imply that realization utility provides a complete description of the forces driving investor behavior, even in the context of our experiment. However, the fact that activity in the decision-making circuitry corresponds to some of the computations hypothesized by the realization utility model provides novel evidence that realization utility plays a significant role in the decisions made by our experimental subjects. It further suggests that mechanisms of this kind might also be at work in the real-world transactions of individual investors.

Using neural data to test an economic model is an unusual exercise in the field of economics because a common view in the profession is that models make as-if predictions about behavior, and are not to be taken as literal descriptions of how decisions are actually made (Gul and Pesendorfer (2008); Bernheim (2009)). In contrast to this view, we adopt a neuroeconomic approach which is based on the idea that knowledge about the computational processes that the brain uses to make decisions should be of central interest to economists because, since these processes describe the actual determinants of observed behavior, they provide valuable insights into the drivers of economic behavior (Camerer et al. (2005); Camerer (2007); Fehr and Rangel (2011)).

⁵ See Blood and Zatorre (2001), De Araujo et al. (2003), Kringelbach et al. (2003), Rolls et al. (2003), Small et al. (2003), McClure et al. (2004), Plassmann et al. (2008).

Our study contributes to the nascent field of neurofinance, which seeks to characterize the computations undertaken by the brain to make financial decisions, and to understand how these computations map to behavior. Several early contributions are worth highlighting. Lo and Repin (2002) investigated the extent to which professional experience affects the emotional arousal of traders in stressful situations, where arousal was measured using skin conductance responses and changes in blood pressure. Kuhnen and Knutson (2005) measured neural responses using fMRI during a simple investment task and found that activity in brain regions previously associated with emotional processing, such as the nucleus accumbens and the insula, predicted subjects' subsequent willingness to take risks. Knutson et al. (2008) took these ideas further by showing that exogenous emotional cues (e.g., erotic pictures) could be used to affect investment behavior, and that these cues increased activity in the same areas that they identified in their previous study. More recently, Bruguier et al. (2010) have shown that neural fMRI measurements of the extent to which subjects activate brain areas associated with concrete cognitive skills, such as the ability to predict others' state of mind, might be useful in identifying which subjects would be successful traders.

Our paper contributes to this literature by showing, for the first time, that a combination of fMRI neural measurements and careful experimental design can be used to test the validity of specific economic theories of financial decision making. Our work also contributes more broadly to the rapidly growing field of neuroeconomics, which seeks to characterize the computations made by the brain in different types of decisions, ranging from simple choices to choices involving risk, self-control and complex social interactions. For recent reviews, see Fehr and Camerer (2007), Glimcher et al. (2008), Rangel et al. (2008), Bossaerts (2009), Kable and Glimcher (2009), Rangel and Hare (2010), and Fehr and Rangel (2011).

The paper is organized as follows. Section I presents some background information about the disposition effect and realization utility. Section II describes the experimental design and the predictions of the realization utility hypothesis. Section III provides a detailed description of how

the neural predictions can be tested using fMRI. Section IV describes the results. Section V briefly concludes.

I. Background: The Disposition Effect and the Realization Utility Model

Using an argument based on Kahneman and Tversky's (1979) prospect theory, Shefrin and Statman (1985) predict that individual investors will have a greater propensity to sell stocks trading at a gain relative to purchase price, rather than stocks trading at a loss. They label this the "disposition effect" and provide some evidence for it using records of investor trading. More detailed evidence for the effect can be found in Odean (1998), who analyzes the trading activity, from 1987 to 1993, of 10,000 households with accounts at a large discount brokerage firm. The phenomenon has now been replicated in several other large databases of trading behavior.

It will be useful to explain Odean's (1998) methodology in more detail because we will adopt a similar methodology in our own analysis. For any day on which an investor in Odean's (1998) sample sells shares of a stock, each stock in his portfolio on that day is placed into one of four categories. A stock is counted as a "realized gain" ("realized loss") if it is sold on that day at a price that is higher (lower) than the average price at which the investor purchased the shares. A stock is counted as a "paper gain" ("paper loss") if its price is higher (lower) than its average purchase price, but it is *not* sold on that day. From the total number of realized gains and paper gains across all accounts over the entire sample, Odean (1998) computes the Proportion of Gains Realized (PGR):

$$PGR = \frac{\# \text{ of realized gains}}{\# \text{ of realized gains} + \# \text{ of paper gains}}$$

In words, PGR computes the number of gains that were realized as a fraction of the total number of gains that could have been realized. A similar ratio, PLR, is computed for losses:

$$PLR = \frac{\# \text{ of realized losses}}{\# \text{ of realized losses} + \# \text{ of paper losses}}$$

The disposition effect is the empirical fact that PGR is significantly greater than PLR. Odean (1998) reports $PGR = 0.148$ and $PLR = 0.098$.

While the disposition effect is a robust empirical phenomenon, its causes remain unclear. This is due, in large part, to the fact that standard rational models of trading have had trouble capturing important features of the data. Consider, for example, an information model in which investors sell stocks with paper gains because they have private information that these stocks will subsequently do poorly, and hold on to stocks with paper losses because they have private information that these stocks will rebound. This hypothesis is inconsistent with Odean's finding that the average return of the prior winners sold by investors is 3.4% *higher*, over the next year, than the average return of the prior losers they hold on to. Another natural model involves taking into account the favorable treatment of losses by the tax code. However, this model also fails to explain the disposition effect because tax-loss selling predicts a greater propensity to sell stocks associated with paper *losses*. Another model attributes the disposition effect to portfolio rebalancing of the kind predicted by a standard framework with power utility preferences and i.i.d. returns. However, under this hypothesis, rebalancing is the "smart" thing to do, which implies that we should observe a stronger disposition effect for more sophisticated investors. In contrast to this prediction, it is *less* sophisticated investors who exhibit a stronger disposition effect (Dhar and Zhu (2006)).

Early on, researchers proposed behavioral economics models of the disposition effect, which can potentially explain the stylized facts that the rational explanations just described cannot explain. One popular model assumes that investors have an irrational belief in mean-reversion (Odean (1998); Weber and Camerer (1998); Kaustia (2010)). If investors *believe* that stocks that have recently done well will subsequently do poorly, and that stocks that have recently done poorly will subsequently do well, their optimal trading strategy would lead to a disposition

effect. We label such beliefs “irrational” because they are at odds with Odean’s (1998) finding that the winner stocks investors sell subsequently do well, not poorly. While the mean-reversion hypothesis is appealing for its simplicity, and is consistent with some evidence from psychology on how people form beliefs⁶, some studies cast doubt on its empirical validity. Weber and Camerer (1998) ask subjects to trade stocks in an experimental stock market, and find that they exhibit a disposition effect in their trading. In order to test the mean-reversion hypothesis, they add a condition in which subjects’ holdings are exogenously liquidated at full value at random times, after which subjects are asked to reinvest the proceeds across stocks in any way they like. Note that if subjects are holding on to stocks with paper losses because of a belief in mean-reversion, we would expect them to re-establish their positions in these stocks, but in fact, they do not.⁷

Another popular behavioral economics model posits that the disposition effect results from prospect theoretic preferences (Kahneman and Tversky (1979)). Prospect theory is a prominent theory of decision-making under risk which assumes that individuals make decisions by computing the utility of potential gains and losses measured relative to a reference point that is often assumed to be the status quo, and that utility is concave over gains and convex over losses. At first sight, it appears that prospect theory preferences may be helpful for understanding the disposition effect. If an investor is holding a stock that has risen in value, he may think of it as trading at a gain. Moreover, if the concavity of the value function over gains induces risk aversion, this may lead him to sell the stock. Conversely, if the convexity of the value function over losses induces risk-seeking, he may be inclined to hold on to a stock that has dropped in value. Contrary to this intuition, Barberis and Xiong (2009) have recently shown that it is surprisingly difficult to derive behavior consistent with the disposition effect using this model. In fact, they show that an investor who derives prospect theory utility from the annual trading profit

⁶ For a review, see Rabin (2002).

⁷ Odean (1998) and Kaustia (2010) provide additional evidence that is inconsistent with the mean-reversion hypothesis.

on each stock that he owns will often exhibit the *opposite* of the disposition effect. Further theoretical arguments against this model have been provided by Kaustia (2010), who has shown that it predicts that investors' propensity to sell a stock depends on the magnitude of the embedded paper gain in a way that is inconsistent with the empirical evidence.

Another behavioral model of the disposition effect is based on the realization utility hypothesis (Shefrin and Statman (1985); Barberis and Xiong (2011)). The central assumption of this model is that investors derive *direct* utility from realizing capital gains and losses on risky assets that they own: they experience a positive burst of utility when they sell an asset at a gain relative to purchase price, where the amount of utility depends on the size of the realized gain; and a negative burst when they sell an asset at a loss relative to purchase price, where the amount of disutility again depends on the size of the loss realized. Importantly, this hypothesis states that trades have a direct utility impact on investors, not just an indirect one through their effect on lifetime wealth and consumption.⁸ Barberis and Xiong (2011) show that linear realization utility, combined with a sufficiently high time discount rate, leads to a disposition effect. The intuition is simple. If an investor derives pleasure from realizing capital gains and, moreover, is impatient, he will be very keen to sell stocks at a gain. Conversely, if he finds it painful to sell stocks at a capital loss and also discounts future utility at a high rate, he will delay selling losing stocks for as long as possible.⁹

⁸ Barberis and Xiong (2011) speculate that realization utility might arise because of the way people think about their investing history. Under this view, some investors – in particular, less sophisticated investors -- do not think about their investing history in terms of overall portfolio return, but rather as a series of investing “episodes,” each of which is characterized by three things: the identity of the asset, the purchase price, and the sale price. “I bought GE at \$40 and sold it at \$70” might be one such episode, for example. According to this view, an investor who sells a stock at a gain feels a burst of positive utility right then because, through the act of selling, he is creating a positive new investing episode. Similarly, if he sells a stock at a loss, he experiences a burst of disutility: by selling, he is creating a negative investing episode.

⁹ Time discounting is not a critical part of the realization utility hypothesis. The disposition effect also follows from realization utility combined with an S-shaped value function, as in prospect theory (Barberis and Xiong, 2009). Adopting this interpretation of the realization utility hypothesis would not significantly affect the analysis that follows.

While the realization utility hypothesis makes predictions about behavior that are consistent with the disposition effect, as well as with other empirical patterns¹⁰, it is based on assumptions that depart significantly from those of traditional models. In particular, its predictions rely on the assumption that utility depends not only on consumption, but also on capital gains and losses realized from the sale of specific assets. Given the unusual nature of this assumption, it seems especially important to carry out direct tests of the extent to which the hypothesized source of utility is actually computed by subjects and affects their decisions. In the rest of the paper we show how this can be done using a combination of fMRI measures of neural activity and careful experimental design.

II. Experimental Design and Predictions

In this section, we first describe the experimental stock market that we set up to test the realization utility model. We then lay out the specific behavioral and neural predictions of the theory that we test.

A. Design

The design of the experimental stock market builds directly on an earlier non-neural experiment conducted by Weber and Camerer (1998).

Subjects are given the opportunity to trade three stocks – stock A, stock B, and stock C – in an experimental market. The experiment consists of two identical sessions separated by a one-

¹⁰ Barberis and Xiong (2011) show that realization utility can shed light on many empirical phenomena, not just on the disposition effect. Some of the other applications they discuss are the poor trading performance of individual investors, the greater volume of trading in bull markets than in bear markets, the individual investor preference for volatile stocks, and the low average return of volatile stocks.

minute break. Each session lasts approximately 16 minutes and consists of 108 trials. We use t to index the trials within a session.¹¹

At the beginning of each session, each subject is given \$350 in experimental currency and is required to buy one share of each stock. The initial share price for each stock is \$100; after the initial purchase, each subject is therefore left with \$50. Every trial $t > 9$ consists of two parts: a price update and a trading decision, each of which corresponds to a separate screen that the subject sees (Figure 1). In the price update part, one of the three stocks is chosen at random and the subject is shown a price change for this stock. Note that stock prices only evolve during the price update screens; as a result, subjects see the entire price path for each stock. In the trading part, one of the three stocks is again chosen at random and the subject is asked whether he wants to trade the stock. Note that no new information is revealed during this part.

We split each trial into two parts so as to temporally separate different computations associated with decision-making. At the price update screen, subjects are provided with information about a change in the price of one of the three stocks, but do not have to compute the value of buying or selling the stock, both because they are not allowed to make decisions at this stage, and also because they do not know which of the three assets will be selected for trading in the next screen. At the trading screen the opposite situation holds: subjects need to compute the value of buying or selling a stock, but do not need to update their beliefs about the price process since no new information about prices is provided.

Trials 1 through 9 consist only of a price update stage; i.e., subjects are not given the opportunity to buy or sell during these trials. We designed the experiment in this way so that subjects can accumulate some information about the three stocks before having to make any trading decisions.

¹¹ We split our experiment into two sessions in order to avoid running the fMRI machine for too long without a break, as this could lead to potential medical risks for the subjects.

Each subject is allowed to hold a maximum of one share and a minimum of zero shares of each stock at any point in time. In particular, short-selling is not allowed. The trading decision is therefore reduced to deciding whether to sell a stock (conditional on holding it), or deciding whether to buy it (conditional on not holding it). The price at which a subject can buy or sell a stock is given by the current market price of the stock.

The price path of each stock is governed by a two-state Markov chain with a good state and a bad state. The Markov chain for each stock is independent of the Markov chains for the other two stocks. Suppose that, in trial t , there is a price update for stock i . If stock i is in the good state at that time, its price increases with probability 0.55 and decreases with probability 0.45. Conversely, if it is in the bad state at that time, its price increases with probability 0.45 and decreases with probability 0.55. The magnitude of the price change is drawn uniformly from $\{\$5, \$10, \$15\}$, independently of the direction of the price change.

The state of each stock changes over time in the following way. Before trial 1, we randomly assign a state to each stock. If the price update in trial $t > 1$ is *not* about stock i , then the state of stock i in trial t remains the same as its state in the previous trial, $t-1$. If the price update in trial $t > 1$ *is* about stock i , then the state of stock i in this trial remains the same as in trial $t-1$ with probability 0.8, but switches with probability 0.2. In mathematical terms, if $s_{i,t} \in \{\text{good}, \text{bad}\}$ is the state of stock i in trial t , then $s_{i,t} = s_{i,t-1}$ if the time t price update is not about stock i , whereas if the time t price update is about stock i , the state switches as follows:

	$s_{i,t+1}=\text{good}$	$s_{i,t+1}=\text{bad}$
$s_{i,t}=\text{good}$	0.8	0.2
$s_{i,t}=\text{bad}$	0.2	0.8

The states of the stocks are never revealed to the subjects: they have to infer them from the observed price paths. To ease comparison of trading performance across subjects, the same set of realized prices is used for all subjects.

A key aspect of our design is that, conditional on the information available to subjects, each of the stocks exhibits positive short-term autocorrelation in its price changes. If a stock performed well on the last price update, it was probably in a good state for that price update. Since it is highly likely (probability 0.8) to remain in the same state for its next price update, its next price change is likely to also be positive.

At the end of each session, we liquidate subjects' holdings of the three stocks and record the cash value of their position. We give subjects a financial incentive to maximize the final value of their portfolio at the end of each session. Specifically, if the total value of a subject's cash and risky asset holdings at the end of session 1 is $\$X$, in experimental currency, and the total value of his cash and risky asset holdings at the end of session 2 is $\$Y$, again in experimental currency, then his take-home pay in *actual* dollars is $15 + (X+Y)/24$.¹² Subjects' earnings ranged from \$43.05 to \$57.33 with a mean of \$52.57 and a standard deviation of \$3.35.

In order to avoid liquidity constraints, we allow subjects to carry a negative cash balance in order to purchase a stock if they do not have sufficient cash to do so at the time of a decision. If a subject ends the experiment with a negative cash balance, this amount is subtracted from the terminal value of his portfolio. The large cash endowment, together with the constraint that subjects can hold at most one unit of each stock at any moment, was sufficient to guarantee that no one ended the experiment with a negative portfolio value, or was unable to buy a stock because of a shortage of cash during the experiment.

N=28 Caltech subjects participated in the experiment (22 male, age range 18 – 60).¹³ All subjects were right-handed and had no history of psychiatric illness, and none were taking medications that interfere with fMRI. The exact instructions given to subjects at the beginning of the experiment are included in the Appendix. The instructions carefully describe the stochastic

¹² In other words, we average X and Y to get $(X+Y)/2$, convert the experimental currency to actual dollars using a 12:1 exchange rate, and add a \$15 show-up fee.

¹³ One additional subject participated in the experiment but was excluded from further analyses because his head motion during the scanning exceeded a pre-specified threshold, thereby interfering with the reliability of the neural measurements.

structure of the price process, as well as all other details of the experiment. Before entering the scanner, the subjects underwent a practice session of 25 trials to ensure familiarity with the market software.

Finally, note that, in our experiment, there is a straightforward way to measure the extent to which a subject exhibits a disposition effect in his trading. We simply adapt Odean's (1998) methodology, described in Section I, in the following way. Every time a subject faces a decision about selling a stock, we classify his eventual action as a paper gain (loss) if the stock's current price is above (below) the purchase price and he chooses not to sell; and as a realized gain (loss) if the stock's current price is above (below) the purchase price and he chooses to sell. We then count up the number of paper gains, paper losses, realized gains, and realized losses over all selling decisions faced by the subject and compute the PGR and PLR measures described earlier. We assign the subject a disposition effect measure of PGR-PLR. When this measure is positive (negative), the subject exhibits (the opposite of) a disposition effect.

B. Optimal trading strategy

We now characterize the optimal trading strategy for a risk-neutral Bayesian investor who is maximizing the expected value of his take-home earnings – from now on, we refer to such an investor as an “expected value” investor. The optimal strategy of such an investor is to sell (or not buy) a stock when he believes that it is more likely to be in the bad state than in the good state; and to buy (or hold) the stock when he believes that it is more likely to be in the good state. Formally, let $p_{i,t}$ be the price of stock i in trial t , after any price update about the stock, and let $q_{i,t} = \Pr(s_{i,t} = \text{good} \mid p_{i,t}, p_{i,t-1}, \dots, p_{i,1})$ be the probability that a Bayesian investor, after seeing the price update in trial t , would assign to stock i being in the good state in trial t . Also, let z_t take the value 1 if the price update in trial t indicates a price increase for the stock in question; and -1 if

the price update indicates a price decrease. Then $q_{i,t} = q_{i,t-1}$ if the price update in trial t was *not* about stock i ; but if the price update in trial t was about stock i , then:

$$(1) \quad q_{i,t}(q_{i,t-1}, z_t) = \frac{\Pr(z_t | s_{i,t} = \text{good}) * \Pr(s_{i,t} = \text{good} | q_{i,t-1})}{q_{i,t-1} \Pr(z_t | s_{i,t-1} = \text{good}) + (1 - q_{i,t-1}) \Pr(z_t | s_{i,t-1} = \text{bad})}$$

$$= \frac{(0.5 + 0.05z_t) * [0.8 * q_{i,t-1} + 0.2 * (1 - q_{i,t-1})]}{q_{i,t-1} [0.8 * (0.5 + 0.05z_t) + 0.2 * (0.5 - 0.05z_t)] + (1 - q_{i,t-1}) [0.2 * (0.5 + 0.05z_t) + 0.8 * (0.5 - 0.05z_t)]}$$

The optimal strategy for an expected value investor is to sell (if holding) or not buy (if not holding) stock i in trial t when $q_{i,t} < 0.5$; and to hold or buy it otherwise.

Note that a trader who follows the optimal strategy described above will exhibit the opposite of the disposition effect. If a stock performed well on the last price update, it was probably in a good state for that price update. Since it is very likely to remain in the same state for its next price update, its next price change is likely to also be positive. The optimal strategy therefore involves selling winner stocks relatively rarely, and losing stocks more often, thereby generating the reverse of the disposition effect.

Of course, it is difficult for subjects to do the exact calculation in equation (1) in real time during the experiment. However, it is relatively straightforward for subjects to approximate the optimal strategy: they need simply keep track of each stock's most recent price changes, and then hold on to stocks that have recently performed well while selling stocks that have recently performed poorly. The fact that a stock's purchase price is reported on the trading screen makes it particularly easy to follow an approximate strategy of this kind: subjects can simply use the difference between the current market price and the purchase price as a proxy for the stock's recent performance.¹⁴

¹⁴ Our rational benchmark assumes risk-neutrality because the monetary risk in our experiment is small. We have also considered the case of risk aversion, however, and have concluded that its predictions do not differ significantly from those of risk neutrality. In some frameworks, risk aversion can generate a disposition effect through rebalancing motives. This is not the case in our experiment, however, because the volatility of stock price changes is independent of the level of the stock price. Furthermore, any

C. Behavioral and neural predictions of the realization utility model

We now lay out the behavioral and neural predictions of the realization utility model, and contrast them with the predictions made by the expected value agent model.

Consider the behavioral predictions first. Since the experimental stock prices exhibit short-term momentum, an expected value investor will exhibit the opposite of the disposition effect: for the actual price process that our subjects see, the value of the PGR-PLR measure of the disposition effect under the optimal trading strategy for an expected value investor is -0.76. In other words, such an investor will have a much greater propensity to realize losses than to realize gains. By contrast, a trader who experiences bursts of realization utility and who discounts future utility at a high rate will sell winner stocks more often than the expected value trader and loser stocks less often. After all, he is keen to realize capital gains as soon as possible and to postpone realizing capital losses as long as possible. This leads to our first prediction.

Prediction 1 (Behavioral): For an expected value investor, the value of the PGR-PLR measure is given by -0.76. On the other hand, for a subject who experiences bursts of realization utility, the value of PGR-PLR is greater than -0.76.

We now turn to the neural predictions made by the two models. As noted earlier, a maintained assumption here is that the theories are not only making predictions about behavior, but are also describing the key computations that subjects have to undertake in order to make decisions.

The first two neural predictions build on a basic finding from the field of decision neuroscience. A sizable number of studies have found evidence consistent with the idea that in

rebalancing motives would be of second-order importance relative to time variation in the *mean* stock return.

simple choice situations the brain makes decisions by assigning values (often called “decision values”) to the options under consideration, and then comparing them to make a choice¹⁵. These value signals are thought to reflect the relative value of taking the action or option under consideration (e.g., sell a stock) versus staying with the status quo (e.g., don’t sell it) (De Martino et al. (2006); Hutcherson et al. (2011)). A significant body of work, using various neural measurement techniques, has shown that activity in regions of the ventromedial prefrontal cortex (vmPFC), and often also the ventral striatum (vSt), correlates with the decision values of options across a range of choices. For example, a recent study shows that, when subjects have to make purchasing decisions for goods such as monetary lotteries, foods, or DVDs, activity in the vmPFC correlates with behavioral measures of their willingness to pay (their “decision value”) taken prior to the choice task (Chib et al. (2009)). See Rangel and Hare (2010) for an up-to-date review of the evidence.

Now consider the decision value signals that would be computed at the time of making a selling decision by an individual who makes choices according to the expected value model. In the context of our experiment, the decision value of selling a stock is given by the value of selling the stock minus the value of holding it. For the expected value investor, the value of selling the stock is zero: if he sells it, he will no longer own any shares of it, and so it can no longer generate any value for him. In contrast, the value of holding the stock can be approximated by the stock’s expected price change on its next price update:

$$E_t[\Delta p_{i,t+1} | q_{i,t}, \Delta p_{i,t+1} \neq 0] = 0.6(2q_{i,t} - 1).$$

It follows that the decision value signal at the time of making a selling decision is given by $0 - 0.6(2q_{i,t} - 1)$, or $0.6(1 - 2q_{i,t})$; we will refer to this quantity throughout the paper as the *net expected value* of selling, or NEV. Note that this is only an approximation because the exact value of

¹⁵ See, for example, Hsu et al. (2005), Padoa-Schioppa and Assad (2006), Kable and Glimcher (2007), Knutson et al. (2007), Tom et al. (2007), Hare et al. (2008), Kennerley et al. (2008), Chib et al. (2009), Hare et al. (2009), Hsu et al. (2009), Hare et al. (2010), Levy et al. (2010), Litt et al. (2010), Rangel and Hare (2010).

holding a stock is the stock's expected *cumulative* price change until the subject decides to sell it. However, this approximation has little effect on our later results because the value of holding a stock until its next price change is highly correlated with the value of holding the stock until it is actually optimal to sell it (the latter quantity can be computed by simulation).

Now consider the decision value signal that would be computed at the time of making a selling decision by an individual who makes choices according to the realization utility model. In particular, consider a simple form of the model in which subjects maximize the sum of expected discounted realized capital gains and losses. For such a trader, the value of selling is linearly proportional to the capital gain or loss, given by $p_{i,t} - c$, where c is the purchase price, or cost basis. However, the expected impact of holding the stock on realization utility is approximately zero, as long as the discount rate is sufficiently high. Thus, for such an investor, the decision value of selling should be linearly related to $p_{i,t} - c$.¹⁶ This, together with the fact that decision value signals have been found to be reliably encoded in the vmPFC and the vSt, leads to the next prediction.

Prediction 2 (Neural): For expected value traders, activity in the regions of the vmPFC and the vSt associated with the computation of decision values should be linearly proportional to the

¹⁶ We say that the value of holding a stock is “approximately” zero for a realization utility investor because, in principle, there is some value to holding, namely expected future realization utility flows. However, under the realization utility hypothesis, the trader is essentially myopic – he discounts future utility flows at a high rate. To a first approximation, then, the value of holding is zero. It may initially seem surprising that a subject would discount future utility at a high rate in the context of a 30-minute experiment. However, the literature on hyperbolic discounting suggests that discounting can be steep even over short intervals, perhaps because people distinguish sharply between rewards available right now, and rewards available at all future times. Furthermore, what may be important in our experiment is not so much calendar time, as transaction time. A subject who can trade stock B now may view the opportunity to trade it in the future as a very distant event -- one that is potentially dozens of screens away – and hence one that he discounts heavily. Finally, we note that discounting is not a critical part of our hypothesis. The disposition effect also follows from a model that combines realization utility with an S-shaped utility function, as in prospect theory. To a first approximation, this model would produce the same decision value as the discounting-based model. The reason is that, under an S-shaped utility function, the utility of selling a stock at a gain (loss) immediately is significantly higher (lower) than the expected utility of holding on to it.

NEV, $0.6(1-2q_{i,t})$, at the time of making selling decisions, and thus independent of the cost basis. In contrast, for subjects who experience realization utility proportional to realized capital gains and losses, activity in these areas of the vmPFC and the vSt should be linearly related to the realizable gain or loss, $p_{i,t} - c$.

The previous arguments predict that traders who place a large weight on realization utility when making decisions should exhibit neural activity in the vmPFC and the vSt that is more strongly correlated with the realizable capital gains or losses. At the same time, subjects with a larger weight on realization utility when making decisions should exhibit a stronger disposition effect. It follows that the degree to which vmPFC and vSt activity correlates with the realizable capital gain should be correlated, across subjects, with the strength of the disposition effect in their trading.

Prediction 3 (Neural): The degree to which vmPFC and vSt activity correlates with the realizable capital gain should be correlated, across subjects, with the strength of the disposition effect in their trading.

The final neural prediction is qualitatively different, in that it seeks to test *directly* if the subject experiences a burst of realization utility at the time of selling a stock that is proportional to the realized capital gain. As before, we can test this prediction using fMRI by building on previous work in neuroscience which has shown that activity in regions of the vSt and the vmPFC correlates reliably with reports of subjective pleasure generated by a wide variety of stimuli – including music, paintings, attractive faces, food, and wine.¹⁷ It follows that, if realizing a

¹⁷ See, for example, Blood and Zatorre (2001), De Araujo et al. (2003), Kringelbach et al. (2003), Rolls et al. (2003), Small et al. (2003), McClure et al. (2004), Plassmann et al. (2008)

capital gain generates a positive burst of experienced utility for the investor, it should increase neural activity in these areas precisely at the moment that the decision is made.

Prediction 4 (Neural): Under the realization utility hypothesis, neural activity in areas known to encode instantaneous experienced utility, such as the vSt or the vmPFC, should increase at the precise moment that individuals decide to realize a capital gain, and decrease at the moment they decide to realize a capital loss.

III. fMRI data collection and analysis

In this section, we provide a primer on how fMRI measures of neural activity are collected and analyzed. For more details, see Huettel et al. (2004), Ashby (2011), and Poldrack et al. (2011).

A. fMRI data collection and measurement

We collected measures of neural activity over the entire brain using BOLD-fMRI, which stands for blood-oxygenated level dependent functional magnetic resonance imaging. BOLD-fMRI measures changes in local magnetic fields that result from local inflows of oxygenated hemoglobin and outflows of de-oxygenated hemoglobin that occur when neurons fire. fMRI provides measures of the BOLD response of relatively small “neighborhoods” of brain tissue known as *voxels*, and is thought to measure the sum of the total amount of neural firing into that voxel as well as the amount of neuronal firing within the voxel.¹⁸

One important complication is that the hemoglobin responses measured by BOLD-fMRI are slower than the associated neuronal responses. Specifically, although the bulk of the neuronal

¹⁸ Note that the neural activity measured by fMRI in a 1-mm³ cube (about the size of a grain of salt) represents the joint activity of between 5,000 to 40,000 neurons, depending on the area of the brain.

response takes place quickly, subsequent BOLD measurements are affected for up to 24 seconds. Figure 2A provides a more detailed illustration of the nature of the BOLD response. In particular, it shows the path of the BOLD signal in response to 1 arbitrary unit of neural activity of infinitesimal duration at time zero. The function plotted here is called the canonical hemodynamic response function (HRF). It is denoted by $h(\tau)$, where τ is the amount of elapsed time since the neural activity impulse, and has been shown to approximate well the pattern of BOLD responses for most subjects, brain areas, and tasks.

Fortunately, the BOLD response has been shown to combine linearly across multiple sources of neural activity (Boynton et al. (1996)). This property, along with a specific functional form of the HRF, allows us to construct a mapping from neural activity to BOLD response so that we can control for BOLD responses that are generated by neural activity over the previous 24 seconds. In particular, if the level of neural activity at any particular time is given by $a(t)$, then the level of BOLD activity at any instant t is well approximated by

$$b(t) = \int_0^\infty h(u)a(t-u)du,$$

which is the convolution between the HRF and the neural inputs. The integral can be interpreted in a straightforward way: it is simply a lagged sum of all the BOLD responses triggered by previous neural activity. This is illustrated in Fig. 2B, which depicts a hypothetical path of neural activity, together with the associated BOLD response.

We acquire two types of MRI data during the experiment in a 3.0 Siemens Tesla Trio MRI scanner with an eight-channel phased array coil. First, we acquire BOLD-fMRI data while the subjects perform the experimental task with a voxel size of 3 mm^3 . We acquire data for the

entire brain (~ 100,000 voxels) every 2.75 seconds.¹⁹ We also acquire high-resolution anatomical scans that we use mainly for realigning the brains across subjects and for localizing the brain activity identified by our analyses.²⁰

B. fMRI data pre-processing

Before the BOLD data can be analyzed to test our hypotheses, it has to be converted into a usable format. This requires the following steps, which are fairly standard – see Huettel et al. (2004), Ashby (2011), & Poldrack et al. (2011) – and were implemented using a specialized but commonly used software package called SPM5 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK).

First, images are corrected for slice acquisition time within each voxel. This is necessary because the scanner does not collect data on all brain voxels simultaneously. This simple step, which involves a non-linear interpolation, temporally realigns the data across all voxels.

Second, we correct for head motion to ensure that the time series of BOLD measurements recorded at a specific spatial location within the scanner was always associated with the same brain location throughout the experiment.²¹

Third, we realign the BOLD responses for each individual into a common neuroanatomical frame (the standard Montreal Neurological Institute EPI template). This step,

¹⁹ More precisely, we acquired gradient echo T2*-weighted echoplanar (EPI) images with BOLD contrast. To optimize functional sensitivity in the orbitofrontal cortex (OFC), a key region of interest, we acquired the images in an oblique orientation of 30° to the anterior commissure–posterior commissure line (Deichmann et al. (2003)). Each volume of images had 45 axial slices. A total of 692 volumes were collected over two sessions. The imaging parameters were as follows: echo time, 30 ms; field of view, 192 mm; in-plane resolution and slice thickness, 3mm; repetition time, 2.75 s.

²⁰ More precisely, we acquired high-resolution T1-weighted structural scans (1 x 1 x 1 mm) for each subject, which were coregistered with their mean EPI images and averaged across subjects to permit anatomical localization of the functional activations at the group level.

²¹ BOLD measurements were corrected for head motion by aligning them to the first full brain scan and normalizing to the Montreal Neurological Institute's EPI template. This entails estimating a six-parameter model of the head motion (3 parameters for center movement, and 3 parameters for rotation) for each volume, and then removing the motion using these parameters. For details, see Friston et al. (1996).

called spatial normalization, is necessary because brains come in different shapes and sizes and, as a result, a given spatial location maps to different brain regions in different subjects. Spatial normalization involves a non-linear re-shaping of the brain to maximize the match with a target template. Although the transformed data are not perfectly aligned across subjects due to remaining neuroanatomical heterogeneity, the process suffices for the purposes of this study. Furthermore, any imperfections in the re-alignment process introduce noise that reduces our ability to detect neural activity of interest.

Fourth, we also spatially smooth the BOLD data for each subject by making BOLD responses for each voxel a weighted sum of the responses in neighboring voxels, with the weights decreasing with distance.²² This step is necessary to make sure that the error structure of the data conforms to the normality assumptions about the error structure of the regression models, described below, that we use to test our hypotheses.

Finally, we remove low-frequency signals that are unlikely to be associated with neuronal responses to individual trials.²³

C. fMRI main data analyses

The key goal of our exercise is to identify regions of the brain, given by collections of spatially contiguous voxels, called *clusters*, where the BOLD response reflects neural activity that implements the computations of interest (e.g., realization utility computations). This is complicated by the fact that, since every voxel contains thousands of neurons, the BOLD responses can be driven by multiple signals. Fortunately, the linear properties of the BOLD signal allow for the identification of the neural signals of interest using standard linear regression methods.

²² Smoothing was performed using an 8 mm full-width half-maximum Gaussian kernel.

²³ Specifically, we applied a high-pass temporal filter to the BOLD data with a cut-off of 128 seconds.

The general procedure is straightforward, and will be familiar to most economists. The analysis begins by specifying two types of variables that might affect the BOLD response: target computations and additional controls. The target computations reflect the signals that we are looking for (e.g., a realization utility signal at the time of selling a stock). They are specified by a time series $s_i(t)$ describing each signal of interest. For each of these signals, let $S_i(t)$ denote the time series that results from convolving the signal $s_i(t)$ with the HRF, as described above. The additional controls, denoted by $c_j(t)$, are other variables that might affect the BOLD time series (e.g., residual head movement or time trends). These are introduced to further clean up the noise inherent in the BOLD signal, but are not explicitly used in any of our tests. The control variables are not convolved with the HRF because they reflect parameters that affect the measured BOLD responses, and not neural activity that triggers a hemodynamic response.²⁴

The linearity of the BOLD signal implies that the level of BOLD activity in any voxel v should be given by

$$b_v(t) = \text{constant} + \sum_i \beta_i^v S_i(t) + \sum_j \alpha_j^v c_j(t) + \varepsilon(t),$$

where $\varepsilon(t)$ denotes AR(1) noise. This model is estimated independently in each of the brain's voxels using standard regression methods.

Our hypotheses can then be restated as tests about the coefficients of this regression model: signal i is said to be associated with activity in voxel v only if β_i^v is significantly different from zero.

Two additional considerations apply to most fMRI studies, including the present one. First, we are interested in testing hypotheses about the distribution of the signal coefficients in the population, and not about individual coefficients. This requires estimating a random effects version of the linear model specified above which, given the size of a typical fMRI dataset, is

²⁴ For example, linear trends are often included because the scanner heats up with continuous operation and this induces a linear change in the measured BOLD responses.

computationally intensive. Fortunately, it has been shown that there is a straightforward shortcut that provides a good approximation to the full mixed effects analysis (Penny et al. (2006)). It involves estimating the parameters separately for each individual subject, averaging them across subjects, and then performing *t*-tests. This is the approach we follow here.

Second, given that these tests are carried out in each of the ~100,000 voxels in the brain, there is a serious concern about false-positives, and multiple comparison corrections are necessary. Several approaches have been proposed in the fMRI literature to address this problem, many of which rely on the idea that purely random activations are unlikely to come in sizable clusters.²⁵ Here, we follow a common approach in the literature, which consists of combining a sizable statistical threshold for the test in each voxel, given by $p < 0.001$ uncorrected, together with a minimal cluster size of 15 voxels. These two criteria, taken together, severely reduce the likelihood of false positives.

The analyses described so far involve searching for neural correlates of signals of interest across the entire brain and are therefore known as whole brain analyses. Another popular and very useful type of exercise, which we use here, is a “region of interest” (ROI) analysis. Put simply, this analysis differs from a whole-brain analysis because it first restricts the set of voxels that is being analyzed. The most common types of ROI analyses involve 1) the measurement of signal strength in a pre-specified ROI (in other words, in a pre-specified subset of voxels), 2) computing the correlation across subjects between measures of signal strength in a particular ROI and behavioral or psychological measures, and 3) characterizing the time course of BOLD responses in an ROI for a particular event (e.g, selling a stock.)

The measurement of signal strength in pre-specified ROIs is a straightforward extension of the whole brain analysis. In this case, a general linear model is estimated *only* for the voxels in the ROI, and then a response estimate for the signal of interest is computed for every subject by

²⁵ As noted earlier, a cluster is a set of spatially contiguous voxels.

averaging over the estimated coefficients over all of the voxels in the ROI. The distribution of average estimates for the group can then be compared across signals of interest using t-tests.

The characterization of the time course of BOLD responses in specific ROIs and around particular events is a little more complicated, but is needed in order to conduct a test of Prediction 4. It requires the specification of a version of the GLM described above that uses a series of “event-locked” dummy variables. The nature of this model is most easily explained with a concrete example. Suppose that we are interested in characterizing the time course of changes in BOLD activity that follows the rapid presentation of two types of images to subjects, type A and B. We then define a series of dummy variables

$$d(t|x, n) = \begin{cases} 1 & \text{if stimulus } x \text{ was presented at } t - n \\ 0 & \text{otherwise} \end{cases}$$

for $x=A, B, n=1, \dots, T$. The general model is then specified as

$$b_v(t) = \text{constant} + \sum_{x,n} \gamma_{x,n} d(t|x, n) + \varepsilon(t).$$

The estimate of the change in the BOLD response n seconds after the presentation of stimulus x is then given by $\gamma_{x,n}$.

IV. Results

A. Behavioral predictions

We begin our test of Prediction 1 by computing the strength of the disposition effect for each subject using the PGR-PLR measure described at the end of Section IIA. We find that the average PGR and PLR across subjects are .412 and .187, respectively. This implies an average

PGR-PLR value of 0.225, which is significantly greater than 0 ($p < 0.001$). In other words, not only is the average value of PGR-PLR significantly greater than the expected value benchmark of -0.76, but it is actually significantly positive. These results reject the hypothesis that our subjects are all expected value investors and are consistent with the idea that some of our subjects are affected by realization utility.

Figure 3 tests the prediction at the individual level. Each bar shows the value of PGR-PLR for a particular subject. The horizontal dashed line near the bottom of the figure marks the -0.76 value of PGR-PLR that an expected value investor would exhibit. The figure shows that *every* subject exhibits a disposition effect greater than -0.76. The hypothesis that the average disposition effect is not different from -0.76 is rejected with a t -statistic of 16.52.

The figure also shows that there is significant heterogeneity in the strength of the disposition effect across subjects: the value of PGR-PLR ranges from -0.41 to 0.83 and has a standard deviation of 0.32. This cross-individual variation is consistent with Dhar and Zhu (2006) who, using data on actual trading decisions, also find significant variation in the strength of the disposition effect across investors. Interestingly, we find that, while each of PGR and PLR varies a good deal across subjects, the two variables have a correlation of only 0.03: subjects who are slow to sell losing stocks are not necessarily also quick to sell winning stocks²⁶. This independence between selling behavior in the gain domain and in the loss domain is also consistent with the empirical findings of Dhar and Zhu (2006).

Figure 4 provides additional insight into subjects' selling behavior by showing, for each of the four types of decisions that a subject could make – decisions to realize a gain, decisions to realize a loss, decisions not to realize a gain, and decisions not to realize a loss -- the fraction of the decisions that are optimal, where “optimal” is defined by the expected value benchmark. For

²⁶ The low correlation between PGR and PLR is not inconsistent with realization utility; it simply suggests that realization utility is not the only factor driving subjects' trading. For example, if our subjects care to varying extents about realization utility but also differ in how much they enjoy trading in general, they may exhibit a near-zero correlation between PGR and PLR: the negative correlation between the two variables induced by realization utility will be offset by the positive correlation induced by the taste for trading.

example, the figure shows that there were a total of 495 occasions in which our subjects realized gains, and that most of these decisions were suboptimal. Given that stocks exhibit short-term price momentum in the experiment, it is generally better to hold on to a stock that has been performing well. This explains why most (77.9%) of subjects' decisions to hold on to winning stocks were optimal, and why most (67.5%) of subjects' decisions to sell winning stocks were suboptimal. Similarly, in the experiment, it is generally better to sell a stock that has been performing poorly. This explains why most (79.2%) of subjects' decisions to sell losing stocks were optimal, while most (80.3%) of their decisions to hold these stocks were suboptimal.

The disposition effect exhibited by our subjects is stronger than that found in empirical studies (Odean (1998); Frazzini (2006)). One possible reason for this is that the current price and the cost basis of a stock are both prominently displayed on the trading screen.²⁷ If, because of realization utility, a subject has a preference for realizing gains and for not realizing losses, the fact that we report the purchase price might make it particularly easy for him to cater to this preference, and hence to exhibit a disposition effect.²⁸

In summary, the behavioral results indicate that all of our subjects exhibit a strong disposition effect, which is inconsistent with the expected value model, but is consistent with the realization utility model.

²⁷ One natural question about our experiment is how much of the realization utility effect that we have found depends on the fact that we display the original purchase price on the trading screen in a highly salient way. It is important to emphasize that it is unlikely that the presence of a realization utility effect depends critically on this aspect of the design. In follow-up work we have carried out behavioral experiments to investigate the impact of the saliency with which the stock purchase price information is displayed (Frydman and Rangel (2011)). We find that eliminating the purchase price from the trading screen diminishes the size of the disposition effect, but that it is still well above the optimal level that an expected value investor would exhibit. This suggests that reporting the purchase price on the trading screen is not a critical aspect of our current design. Moreover, given that most investors in naturally occurring financial markets have at least a rough sense of the price at which they purchased a stock, displaying the cost basis on the trading screen is likely a better approximation of reality.

²⁸ At the same time, because our experimental design induces a negative correlation between the capital gain and the NEV of selling ($r = -0.55$), the fact that we report the purchase price also makes it easy for an expected value subject to trade in a way that is close to *his* optimal strategy, namely to hold a stock when it has a capital gain and to sell it when it has a capital loss. If a subject is an expected value investor, then, we do not think that presenting the purchase price on the trading screen should bias him towards exhibiting a disposition effect.

B. Neural Prediction 2

We now turn to Prediction 2, which states that, for individuals who experience realization utility at the time of selling assets, activity in areas of the brain associated with the computation of decision values, such as the vmPFC and the vSt, should be correlated with the capital gain variable $(p_t - c_t)$. By contrast, it states that, for expected value subjects, activity in these areas should correlate with the NEV variable, but not with the capital gain.

We test this hypothesis in two stages. First, we estimate the following general linear model (GLM) of BOLD activity for each individual:

$$b_v(t) = \text{constant} + \beta_1^v I_{dec}(t)(p_t - c_t) + \beta_2^v I_{dec}(t)NEV_t + \beta_3^v \text{controls} + \varepsilon(t).$$

Here, $b_v(t)$ denotes BOLD signal at time t in voxel v . $I_{dec,t}$ is an indicator function that equals one at the time when the subject is presented with the opportunity to *sell* a stock at time t . NEV_t denotes the net expected value from selling the stock being considered at time t , namely $0.6(1 - 2q_{i,t})$, and $(p_t - c_t)$ is the realizable capital gain. Finally, the *controls* vector includes regressors that control for physical movement inside the scanner, session-specific effects, and any changes in neural activity that might be due to information processing during the price update screens, which is not an activation of interest for the hypothesis being tested. As described in Section III, the regressors involving computations of interest (here, the non-constant regressors NEV and p-c) are convolved with the HRF²⁹. Finally, inferences about the extent to which the signals of interest are encoded in a given voxel are made by carrying out a one-sided t -test of the individually estimated coefficients (i.e., β_1^v and β_2^v) against zero.

²⁹ The amount of the price change during the price update screen, which represents our control for information processing, is convolved with the HRF because this will generate a BOLD response. Controls for physical movement inside the scanner and session-specific effects are not convolved because they do not elicit a BOLD response.

Although we can carry out these tests in all of the brain's voxels, here we limit our search to voxels that belong to pre-specified anatomical areas of the vmPFC and the vSt. These areas were identified using the AAL digital atlas of the human brain (Tzourio-Mazoyer et al. (2002)). Note that these restrictions make our significance threshold of $p < .001$ uncorrected, together with a minimum cluster size of 15 voxels, even less likely to generate false positives than in the standard whole brain analyses to which it is typically applied.

The results from these tests are consistent with the implications of realization utility noted in Prediction 2: we find a cluster of 67 voxels in the vmPFC where $\beta_1^v > 0$. However, no voxels within the vSt exhibit a correlation with the capital gain at our statistical threshold. The location of the vmPFC voxels is depicted in Figure 5. In contrast, there are no clusters that significantly relate to the NEV variable at our statistical threshold. In short, the neural data is consistent with subjects computing the decision value predicted by realization utility, rather than the decision value predicted by the expected value agent model.

Because of the high correlation between the NEV variable and the capital gain variable ($r = -0.55$), we run a robustness check to make sure that the above results are not driven by spurious collinearity issues. This is done by introducing a single change to the GLM: the capital gain variable is orthogonalized (prior to convolution) to the NEV variable, using a standard Gram-Schmidt algorithm (Strang (1988)). Note that this provides an even more stringent test of the realization utility hypothesis because any shared variance between the two variables is now allocated to the NEV. As before, we find a cluster of 67 voxels in vmPFC that satisfies the significance criterion described above.

We also carry out an ROI analysis designed to test the properties of the vmPFC realization utility signals further. The relevant ROI (i.e., the relevant subset of voxels within the vmPFC) is defined by estimating the simpler GLM,

$$b_v(t) = \text{constant} + \beta_1^v I_{dec}(t)(p_t - c_t) + \beta_2^v \text{controls} + \varepsilon(t),$$

and identifying clusters in the vmPFC that are significantly responsive to the capital gain regressor. For the rest of the paper we refer to the resulting ROI, which contains 154 voxels, as the $\text{vmPFC}_{\text{ROI}}$. Note that we define this ROI using this additional regression to side-step the estimation noise introduced by the high correlation between the capital gain and the NEV regressors.

We then carry out the ROI analysis by estimating the following GLM for each voxel in the newly defined ROI:

$$b_v(t) = \text{constant} + \beta_1^v I_{dec}(t) p_t + \beta_2^v I_{dec}(t) c_t + \beta_3^v \text{controls} + \varepsilon(t).$$

This model is interesting because it allows us to compare the strength of the average beta value in the ROI *separately* for the price and cost basis components of the capital gain. Within $\text{vmPFC}_{\text{ROI}}$, $\beta_1=0.025$ ($p<0.001$) and $\beta_2=-0.023$ ($p<0.01$) and the absolute values of the two coefficients are not significantly different ($p=0.79$). These results demonstrate that the correlation with capital gains that we found above is affected by both the price and cost basis components of the capital gain.

Finally, we carry out a similar ROI analysis to test if the strength of the capital gain signal in $\text{vmPFC}_{\text{ROI}}$ is of similar magnitude in capital gain and capital loss trials. The associated GLM is:

$$b_v(t) = \text{constant} + \beta_1^v I_{dec}(t) I_{\text{cap.gain}} (p_t - c_t) + \beta_2^v I_{dec}(t) I_{\text{cap.loss}} (p_t - c_t) + \beta_3^v \text{controls} + \varepsilon(t),$$

where $I_{\text{cap.gain}}$ and $I_{\text{cap.loss}}$ are indicator variables for trials involving capital gains and capital losses, respectively. The average values of β_1 and β_2 are not significantly different ($p=0.69$).

C. Neural Prediction 3

We now test Prediction 3. Specifically, we check whether, as predicted by the realization utility hypothesis, subjects whose neural activity in the vmPFC at the time of a sell decision is particularly sensitive to the realizable capital gain exhibit a stronger disposition effect.

For every subject, we compute the maximum beta value within the vmPFC_{ROI} for the capital gain and capital loss regressors³⁰. Consistent with Prediction 3, we find that the correlation between β_1 and PGR is 0.78 ($p < 0.001$), indicating that subjects who exhibit stronger vmPFC activation in response to a capital gain *do* have a greater propensity to realize gains. Figure 6, which is a scatterplot of PGR against β_1 , illustrates this graphically.

In contrast, we do not find a significant correlation between β_2 and PLR ($p = 0.18$). One potential post-hoc explanation is that there may be different physiological systems involved in making decisions that involve capital gains and capital losses. Consistent with this hypothesis, the cross-subject correlation between the vmPFC (maximum) sensitivities to capital gains and losses, β_1 and β_2 , is only -0.01.

D. Neural Prediction 4

We now test Prediction 4, which constitutes the most direct test of the realization utility hypothesis, and the one that, in our view, showcases the value of the neural data most clearly. The realization utility hypothesis posits that people experience a positive (negative) hedonic impact when they sell a stock at a gain (loss). Since earlier research in neuroscience suggests that activity in the areas of the vSt and the vmPFC correlate with such measures of experienced utility, or hedonics, we can test the hypothesis by looking at changes in the activity in these two areas at the

³⁰ We use a maximum statistic instead of the average statistic because vmPFC_{ROI} is relatively large (154 voxels) and because of the heterogeneity in anatomical and functional structure of vmPFC across subjects. Since we are using this beta value to test for a correlation (instead of testing for a particular value of the mean), using the max statistic will not bias our results.

moment that a subject decides to sell a stock, and compare it to changes in the activity in these areas at the moment that a subject issues a command to hold a stock.

More concretely, we test the hypothesis by carrying out the following ROI analysis in specific sub-areas of the vSt and vmPFC that have been shown, in previous studies, to correlate with experienced utility. In particular, we define $\text{vmPFC}_{\text{EU-ROI}}$ as the set of voxels that are within 5mm of the voxel whose activity exhibited the highest correlation with experienced utility during consumption of wine in Plassman et al. (2008). Similarly, we define $\text{vSt}_{\text{EU-ROI}}$ as the set of voxels that are within 6mm of the two voxels (bilateral) that exhibited the highest correlation with stimulus salience in Zink et al. (2003). Note that the EU subscript in $\text{vmPFC}_{\text{EU-ROI}}$ emphasizes that this is a different ROI from the one described above in the analysis of decision values, as it involves a different area of the vmPFC, one that has been previously shown to correlate with hedonics.

The ROI analysis involves estimating the time course of responses in these two ROIs during sell trials involving a capital gain, as a function of whether or not the subject chooses to sell. Figure 7A depicts the results of the analysis for the $\text{vSt}_{\text{EU-ROI}}$. The red line indicates changes in the vSt BOLD response for trials in which subjects choose to sell; the blue line shows activity in trials in which subjects choose not to sell. Note that $t=0$ corresponds to the time at which subjects indicate their decision by pressing a button on a hand-held button box-- it is not the time at which the trading screen becomes visible. Interestingly, the figure shows there is no significant difference between the two time series until a decision is made. Afterwards, and consistent with the realization utility hypothesis, activity in the vSt is significantly larger for the next six seconds. The average capital gain for stocks that are held is not significantly different from the average capital gain for stocks that are sold (\$15.77 vs. \$18.35). The effect in Figure 7A is therefore not due to the fact that the stocks subjects sell have larger capital gains than the stocks they hold ($p=0.09$).

Contrary to our expectations, we did not find a similar result in the vmPFC_{EU-ROI} (Figure 7B). One possible explanation is that there might be more heterogeneity across subjects in the anatomical and functional structure of the vmPFC, than in the organization of the vSt, which would perhaps mean that the region identified as vmPFC_{EU-ROI} does not really reflect the areas where experienced utility is computed in our sample. We emphasize, however, that this is pure speculation.

E. Tests of the mean-reversion theory of the disposition effect

As discussed in Section I, one prominent alternative behavioral theory of the disposition effect is that investors believe that stock prices mean-revert (Odean (1998); Weber and Camerer (1998); Kaustia (2010)). In our setting, such a belief would be irrational: stock prices in our experiment exhibit short-term *positive* autocorrelation. Nonetheless, if our subjects, for some reason, *think* that the stock prices in our experiment are mean-reverting, this could potentially explain why they tend to sell stocks that have recently gone up while holding stocks that have recently gone down.

To investigate whether a belief in mean-reversion could be driving our subjects' behavior, we estimate the following mixed effects logistic model to test whether recent price changes can significantly predict subjects' decisions to sell or hold a stock:

$$sell_{i,t,s} = (\alpha + a_i) + \beta_1 NEV_{t,s} + \beta_2 (p_{t,s} - c_{t,s}) + \beta_3 \Delta^1 p_{t,s} + \beta_4 \Delta^2 p_{t,s} + \varepsilon_t \quad (6)$$

where $sell_{i,t,s} = 1$ if subject i sold stock s at time t and 0 if he held it, a_i denotes a subject-level fixed effect, and $\Delta^m p_{t,s}$ denotes the m^{th} most recent price change for stock s (these price changes may not have occurred in consecutive trials because price updates in the experiment take place at

random times). We find that the capital gain is a significant predictor of the propensity to sell, ($t\text{-stat}=10.04$), but that none of the other variables are. In particular, neither β_3 nor β_4 is significantly different from zero ($p=.164$ and $p=0.160$, respectively). In other words, contrary to the mean-reversion hypothesis, recent price changes do not significantly predict the decision to sell.

The neural data can also be used to test some aspects of the mean-reversion hypothesis. In particular, we test if neural activity in either the vmPFC or the vSt is correlated with recent price changes. This is done by estimating the following GLM:

$$b_v(t) = \text{constant} + I_{dec}(t)[\beta_1^v(p_t - c_t) + \beta_2^v\Delta^1 p_{t,s} + \beta_3^v\Delta^2 p_{t,s}] + \beta_4^v \text{controls} + \varepsilon(t)$$

Under the mean reversion hypothesis, the decision value of selling should be positively correlated with recent price changes because a recent price increase indicates a lower expected future return, leading to a higher decision value of selling. Neural activity in the vmPFC and vSt should therefore correlate positively with past price changes. Contrary to this hypothesis, we do not find any activity in the vmPFC that is significantly associated with these regressors.

In summary, then, both the behavioral and the neural analyses cast doubt on the mean-reversion hypothesis.

V. Final Remarks

In this paper, we show that neural data obtained through fMRI techniques can be useful in testing theories of investor behavior. Specifically, we use neural data gathered from subjects trading stocks in an experimental market to test a “realization utility” theory of investor trading. While this theory can potentially explain the disposition effect as well as several other financial phenomena, it relies on an unusual assumption: that people derive utility directly from realizing

gains. We identify the neural predictions of realization utility and find broad support for them in our data. Perhaps most striking of all, we find that, at the moment a subject issues a command to sell a stock at a gain, there is a sharp rise in activity in the ventral striatum, an area of the brain that, based on recent research in cognitive neuroscience, is known to encode feelings of subjective pleasure.

We emphasize that the methods we present in this paper are hardly a substitute for traditional empirical methods in finance. On the contrary, brain imaging techniques are simply complementary tools that can be used to test assumptions about investor behavior that are difficult to evaluate using field data or experimental data alone. In particular, we see neural data as a valuable resource when studying the more psychological dimensions of individual investor behavior, precisely because these may derive from variables that are only observable at the neural level.

REFERENCES

- Ashby, F. Gregory, 2011. *Statistical analysis of fmri data* (The MIT Press).
- Barber, Brad M., and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finance* 55, 773-806.
- Barber, Brad M., and Terrance Odean, 2001, Boys will be boys: Gender, overconfidence, and common stock investment, *Quarterly Journal of Economics* 116, 261-292.
- Barberis, Nicholas, and Richard Thaler, 2003, A survey of behavioral finance, in *Handbook of the economics of finance* (Elsevier).
- Barberis, Nicholas, and Wei Xiong, 2009, What drives the disposition effect? An analysis of a long-standing preference-based explanation, *Journal of Finance* 64, 751-784.
- Barberis, Nicholas, and Wei Xiong, 2011, Realization utility, *Journal of Financial Economics* Forthcoming.
- Bernheim, B. Douglas, 2009, On the potential of neuroeconomics: A critical (but hopeful) appraisal, *American Economic Journal: Microeconomics* 1, 1-41.
- Blood, Anne J., and Robert J. Zatorre, 2001, Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion, *Proceedings of the National Academy of Sciences of the United States of America* 98, 11818-11823.
- Bossaerts, Peter, 2009, What decision neuroscience teaches us about financial decision making, *Annual Review of Financial Economics* 1, 383-404.
- Bossaerts, Peter, and Charles Plott, 2004, Basic principles of asset pricing theory: Evidence from large-scale experimental financial markets, *Review of Finance* 8, 135-169.
- Bossaerts, Peter, Charles Plott, and William R. Zame, 2007, Prices and portfolio choices in financial markets: Theory, econometrics, experiments, *Econometrica* 75, 993-1038.
- Boynton, Geoffrey M., Stephen A. Engel, Gary H. Glover, and David J. Heeger, 1996, Linear systems analysis of functional magnetic resonance imaging in human v1, *The Journal of Neuroscience* 16, 4207-4221.
- Bruguier, Antoine J., Steven R. Quartz, and Peter Bossaerts, 2010, Exploring the nature of "trader intuition", *Journal of Finance* 65, 1703-1723.
- Camerer, Colin F., 2007, Neuroeconomics: Using neuroscience to make economic predictions, *The Economic Journal* 117, C26-C42.
- Camerer, Colin, George Loewenstein, and Drazen Prelec, 2005, Neuroeconomics: How neuroscience can inform economics, *Journal of Economic Literature* 43, 9-64.
- Camerer, Colin, and Keith Weigelt, 1991, Information mirages in experimental asset markets, *Journal of Business* 64, 463-93.
- Camerer, Colin, and Keith Weigelt, 1993, Convergence in experimental double auctions for stochastically lived assets, in Daniel Friedman, and John Rust,

- eds.: *The double auction market: Theories, institutions and experimental evaluations* (Addison-Wesley, Redwood City, CA).
- Campbell, John, 2006, Household finance, *Journal of Finance* 61, 1553-1604.
- Chib, V.S., A. Rangel, S. Shimojo, and J. O'Doherty, 2009, Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex, *Journal of Neuroscience* 29, 12315-12320.
- Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick, 2009, Reinforcement learning and savings behavior, *Journal of Finance* 64, 2515-2534.
- De Araujo, Ivan E. T., Edmund T. Rolls, Morten L. Kringelbach, Francis McGlone, and Nicola Phillips, 2003, Taste-olfactory convergence, and the representation of the pleasantness of flavour, in the human brain, *European Journal of Neuroscience* 18, 2059-2068.
- De Martino, Benedetto, Dharshan Kumaran, Ben Seymour, and Raymond J. Dolan, 2006, Frames, biases, and rational decision-making in the human brain, *Science* 313, 684-687.
- Deichmann, R., J. A. Gottfried, C. Hutton, and R. Turner, 2003, Optimized epi for fmri studies of the orbitofrontal cortex, *NeuroImage* 19, 430-441.
- Dhar, Ravi, and Ning Zhu, 2006, Up close and personal: Investor sophistication and the disposition effect, *Management Science* 52, 726-740.
- Fehr, Ernst, and Colin F. Camerer, 2007, Social neuroeconomics: The neural circuitry of social preferences, *Trends in Cognitive Sciences* 11, 419-427.
- Fehr, Ernst, and Antonio Rangel, 2011, Neuroeconomics for economists: Why economists should care about what the brain does, *Working Paper*.
- Feng, Lei, and Mark S. Seasholes, 2005, Do investor sophistication and trading experience eliminate behavioral biases in financial markets?, *Review of Finance* 9, 305-351.
- Frazzini, Andrea, 2006, The disposition effect and underreaction to news, *Journal of Finance* 61, 2017-2046.
- Frydman, Cary, and Antonio Rangel, 2011, Attention to realization utility modulates the disposition effect, *Caltech Working Paper*.
- Genesove, David, and Christopher Mayer, 2001, Loss aversion and seller behavior: Evidence from the housing market, *Quarterly Journal of Economics* 116, 1233-1260.
- Glimcher, Paul W., Ernst Fehr, Colin Camerer, and Russell A. Poldrack, 2008. *Neuroeconomics: Decision making and the brain* (Academic Press, Amsterdam).
- Grinblatt, Mark, and Matti Keloharju, 2001, What makes investors trade?, *Journal of Finance* 56, 589-616.
- Grinblatt, Mark, and Matti Keloharju, 2009, Sensation seeking, overconfidence, and trading activity, *Journal of Finance* 64, 549-578.
- Gul, Farak, and Wolfgang Pesendorfer, 2008, The case for mindless economics, in Andrew Caplin, and Andrew Shotter, eds.: *The foundations of positive and normative economics* (Oxford University Press).
- Hare, Todd, Colin F. Camerer, Daniel Knöpfle, John O'Doherty, and Antonio Rangel, 2010, Value computations in ventral medial prefrontal cortex during

- charitable decision making incorporate input from regions involved in social cognition, *Journal of Neuroscience* 30, 583-590.
- Hare, Todd, Colin F. Camerer, and Antonio Rangel, 2009, Self-control in decision-making involves modulation of the vmPFC valuation system, *Science* 324, 646-648.
- Hare, Todd, John O'Doherty, Colin F. Camerer, Wolfram Schultz, and Antonio Rangel, 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.
- Hsu, Ming, Meghana Bhatt, Ralph Adolphs, Daniel Tranel, and Colin F. Camerer, 2005, Neural systems responding to degrees of uncertainty in human decision-making, *Science* 310, 1680-1683.
- Hsu, Ming, Ian Krajbich, Chen Zhao, and Colin F. Camerer, 2009, Neural response to reward anticipation under risk is nonlinear in probabilities, *The Journal of Neuroscience* 29, 2231-2237.
- Huettel, Scott, Allen Song, and Gregory McCarthy, 2004. *Functional magnetic resonance imaging* (Sinauer Associates).
- Hutcherson, C.A., B. Bushong, M. Rabin, and Antonio Rangel, 2011, Pro-social behavior is sustained by dissociable neural mechanisms of different effectiveness, *Working paper*.
- Jin, Li, and Anna Scherbina, 2011, Inheriting losers, *Review of Financial Studies* 24.
- Kable, Joseph W., and Paul W. Glimcher, 2007, The neural correlates of subjective value during intertemporal choice, *Nat Neurosci* 10, 1625-1633.
- Kable, Joseph W., and Paul W. Glimcher, 2009, The neurobiology of decision: Consensus and controversy, *Neuron* 63, 733-745.
- Kahneman, Daniel, and Amos Tversky, 1979, Prospect theory: An analysis of decision under risk, *Econometrica* 47, 263-291.
- Kang, Min Jeong, Ming Hsu, Ian M. Krajbich, George Loewenstein, Samuel M. McClure, Joseph Tao-yi Wang, and Colin F. Camerer, 2009, The wick in the candle of learning, *Psychological Science* 20, 963-973.
- Kang, Min Jeong, Antonio Rangel, Mickael Camus, and Colin F. Camerer, 2011, Hypothetical and real choice differentially activate common valuation areas, *The Journal of Neuroscience* 31, 461-468.
- Kaustia, Markku, 2010, Prospect theory and the disposition effect, *Journal of Financial and Quantitative Analysis* 45, 1-36.
- Kennerley, Steven W., Aspandiar F. Dahmubed, Antonio H. Lara, and Jonathan D. Wallis, 2008, Neurons in the frontal lobe encode the value of multiple decision variables, *Journal of Cognitive Neuroscience* 21, 1162-1178.
- Knutson, Brian, Scott Rick, G. Elliott Wimmer, Drazen Prelec, and George Loewenstein, 2007, Neural predictors of purchases, *Neuron* 53, 147-156.
- Knutson, Brian, G. Elliott Wimmer, Camelia M. Kuhnen, and Piotr Winkielman, 2008, Nucleus accumbens activation mediates the influence of reward cues on financial risk taking, *NeuroReport* 19, 509-513.
- Kringelbach, M. L., J. O'Doherty, E. T. Rolls, and C. Andrews, 2003, Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness, *Cerebral Cortex* 13, 1064-1071.

- Kuhnen, Camelia M., and Brian Knutson, 2005, The neural basis of financial risk taking, *Neuron* 47, 763-770.
- Levy, Ifat, Jason Snell, Amy J. Nelson, Aldo Rustichini, and Paul W. Glimcher, 2010, Neural representation of subjective value under risk and ambiguity, *Journal of Neurophysiology* 103, 1036-1047.
- Litt, Ab, Hilke Plassmann, Baba Shiv, and Antonio Rangel, 2010, Dissociating valuation and saliency signals during decision-making, *Cerebral Cortex*.
- Lo, Andrew W., and Dmitry V. Repin, 2002, The psychophysiology of real-time financial risk processing, *Journal of Cognitive Neuroscience* 14, 323-339.
- McClure, Samuel M., Jian Li, Damon Tomlin, Kim S. Cypert, Latané M. Montague, and P. Read Montague, 2004, Neural correlates of behavioral preference for culturally familiar drinks, *Neuron* 44, 379-387.
- Odean, Terrance, 1998, Are investors reluctant to realize their losses?, *Journal of Finance* 53, 1775-1798.
- Padoa-Schioppa, Camillo, and John A. Assad, 2006, Neurons in the orbitofrontal cortex encode economic value, *Nature* 441, 223-226.
- Penny, William, Karl Friston, John Ashburner, Stefan Kiebel, and Thomas Nichols, 2006. *Statistical parametric mapping: The analysis of functional brain images* (Academic Press).
- Plassmann, Hilke, John O'Doherty, Baba Shiv, and Antonio Rangel, 2008, Marketing actions can modulate neural representations of experienced pleasantness, *Proceedings of the National Academy of Sciences* 105, 1050-1054.
- Plott, Charles R., and Shyam Sunder, 1988, Rational expectations and the aggregation of diverse information in laboratory security markets, *Econometrica* 56, 1085-1118.
- Poldrack, Russell A., Jeanette Mumford, and Thomas Nichols, 2011. *Handbook of functional mri data analysis* (Cambridge University Press).
- Rabin, Matthew, 2002, Inference by believers in the law of small numbers, *Quarterly Journal of Economics* 117, 775-816.
- Rangel, Antonio, Colin Camerer, and P. Read Montague, 2008, A framework for studying the neurobiology of value-based decision making, *Nat Rev Neurosci* 9, 545-556.
- Rangel, Antonio, and Todd Hare, 2010, Neural computations associated with goal-directed choice, *Current Opinion in Neurobiology* 20, 262-270.
- Rolls, Edmund T., Morten L. Kringelbach, and Ivan E. T. De Araujo, 2003, Different representations of pleasant and unpleasant odours in the human brain, *European Journal of Neuroscience* 18, 695-703.
- Shefrin, Hersch, and Meir Statman, 1985, The disposition to sell winners too early and ride losers too long: Theory and evidence, *Journal of Finance* 40, 777-790.
- Small, Dana M., Michael D. Gregory, Y. Erica Mak, Darren Gitelman, M. Marsel Mesulam, and Todd Parrish, 2003, Dissociation of neural representation of intensity and affective valuation in human gustation, *Neuron* 39, 701-711.
- Strang, Gilbert, 1988. *Linear algebra and its applications* (Harcourth Brace Jovanovich).

- Tom, Sabrina M., Craig R. Fox, Christopher Trepel, and Russell A. Poldrack, 2007, The neural basis of loss aversion in decision-making under risk, *Science* 315, 515-518.
- Tzourio-Mazoyer, N., B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, 2002, Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain, *NeuroImage* 15, 273-289.
- Weber, Martin, and Colin F. Camerer, 1998, The disposition effect in securities trading: An experimental analysis, *Journal of Economic Behavior & Organization* 33, 167-184.
- Zink, Caroline F., Giuseppe Pagnoni, Megan E. Martin, Mukeshwar Dhamala, and Gregory S. Berns, 2003, Human striatal response to salient nonrewarding stimuli, *The Journal of Neuroscience* 23, 8092-8097.

Figure 1. **Sample screens from a typical trial in the fMRI experiment.** Subjects saw the *price update* screen for two seconds, followed by the *trading* screen for which they had up to three seconds to enter a decision (a blank screen was displayed in between in order to temporally separate neural activity associated with decision-making.) The screens shown below are for a trial in which the subject owns a unit of both stocks A and B. The screens were displayed while subjects were inside the fMRI scanner, and decisions were entered with a handheld device.



Figure 2. **BOLD measurements of neural activity.** (A) Canonical hemodynamic response function that approximates the BOLD response that follows one arbitrary unit of instantaneous neural activity at time 0. (B) Example of a path of neural activity together with the associated BOLD response.

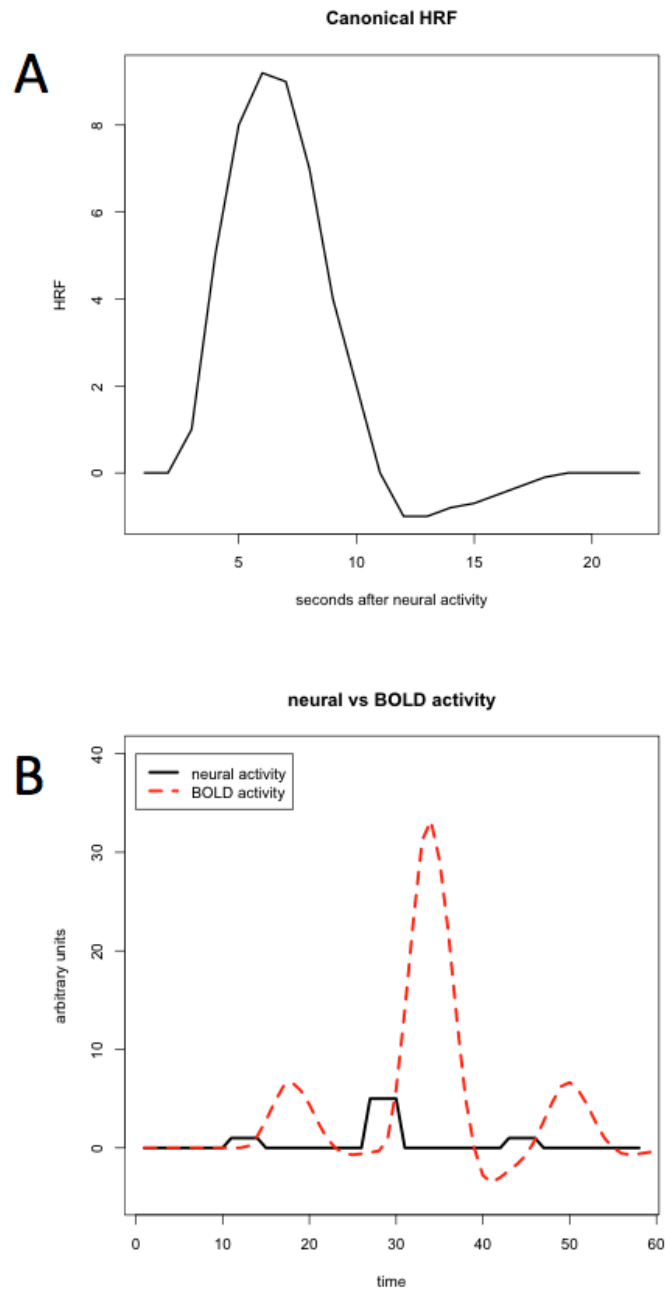


Figure 3. **Measures of the disposition effect (PGR-PLR) for each subject.** Standard error bars are computed as in Odean (1998) and the dotted line indicates the optimal level of the disposition effect, namely -0.76. All subjects exhibit a disposition effect greater than the optimal level and a majority of subjects have a disposition effect that is significantly *positive*. The figure indicates that there is significant heterogeneity in the size of the disposition effect across subjects (SD: 0.32).

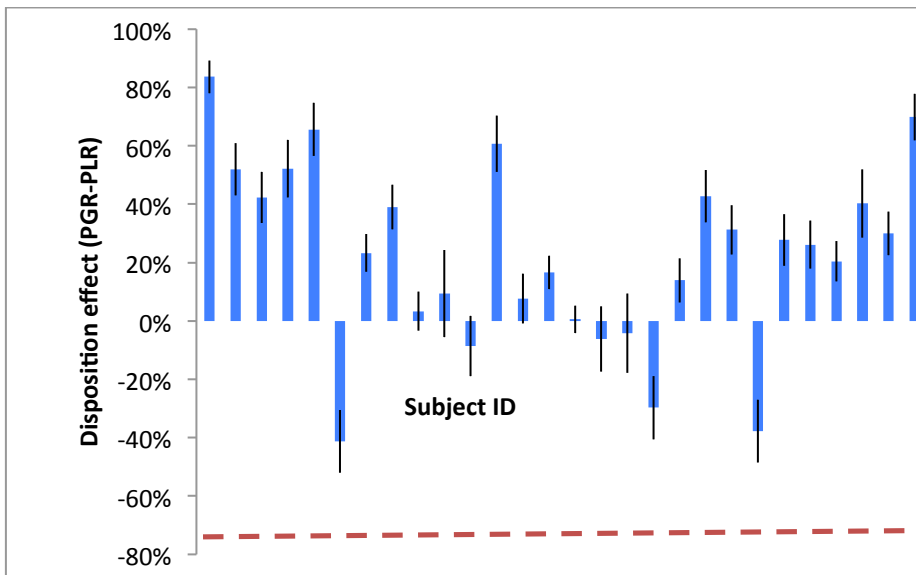


Figure 4. **Total number of sell decisions by decision type and optimality.** Realized gains and losses refer to decisions where subjects sold a stock trading at a gain (loss.) Paper gains (losses) refer to decisions where subjects decided to hold a stock trading at a gain (loss). The optimality measures show an important aspect of our design: selling winners and holding losers, which leads to a disposition effect, are typically suboptimal decisions. Decisions are pooled across all subjects.

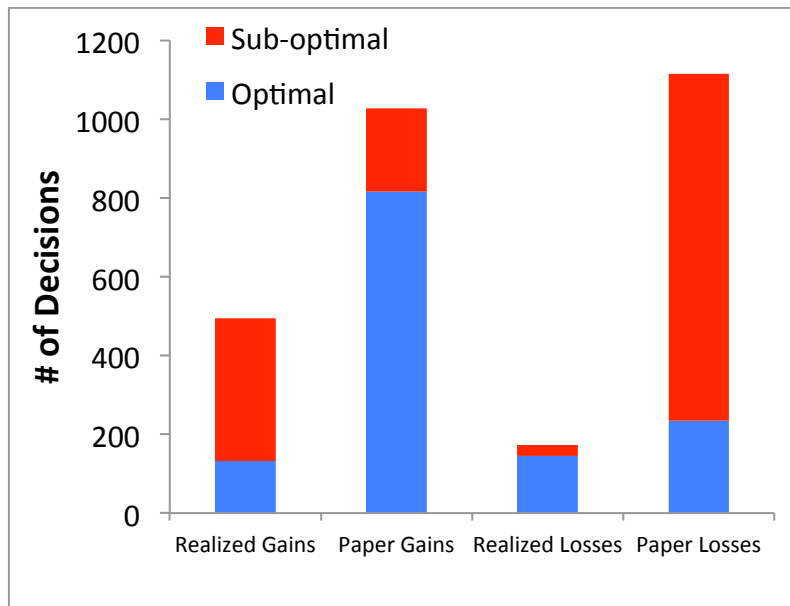


Figure 5. **vmPFC activity reflects realization utility.** Voxels that are shown in yellow all have a p-value less than 0.001, and only clusters of at least 15 significant voxels are shown. Color bar denotes t-statistics.

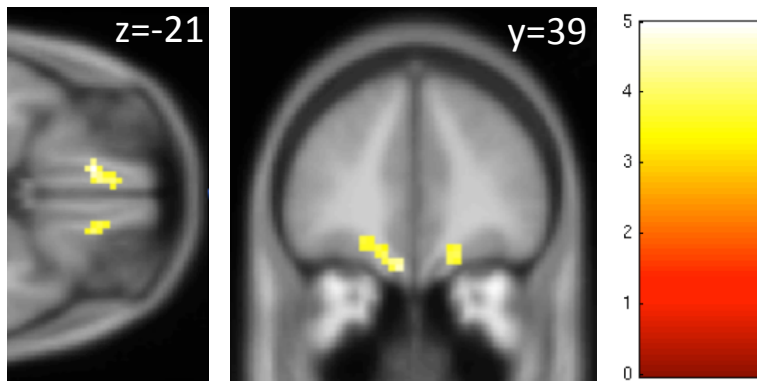
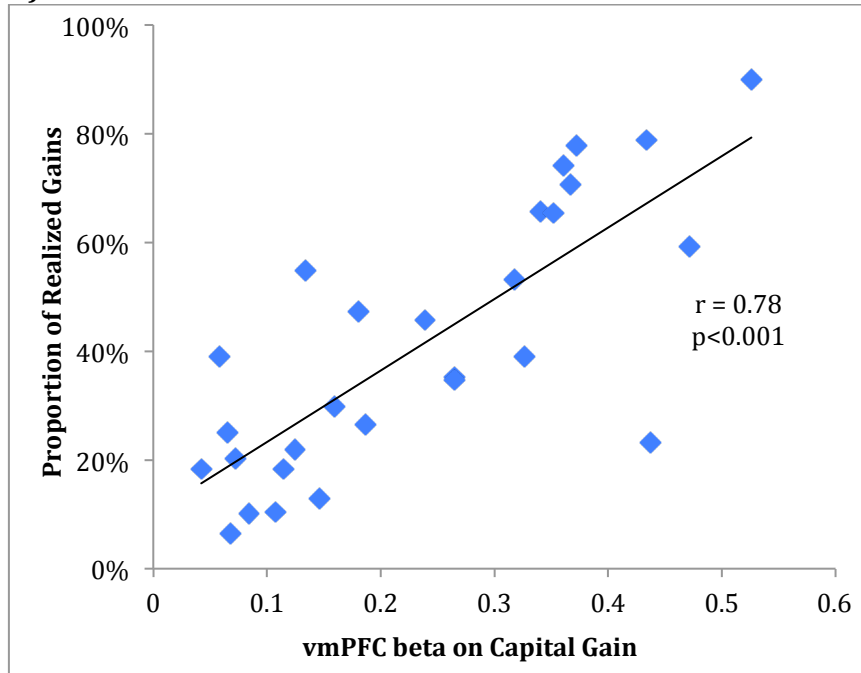


Figure 6. **Correlation between brain activity and measures of the disposition effect.** Each data point in the figure represents a single subject. We find that activity in the vmPFC at the time subjects are offered the opportunity to sell a capital gain is highly correlated with their propensity to realize gains. We do not find a similar correlation between vmPFC activity and the propensity to realize losses.

A)



B)

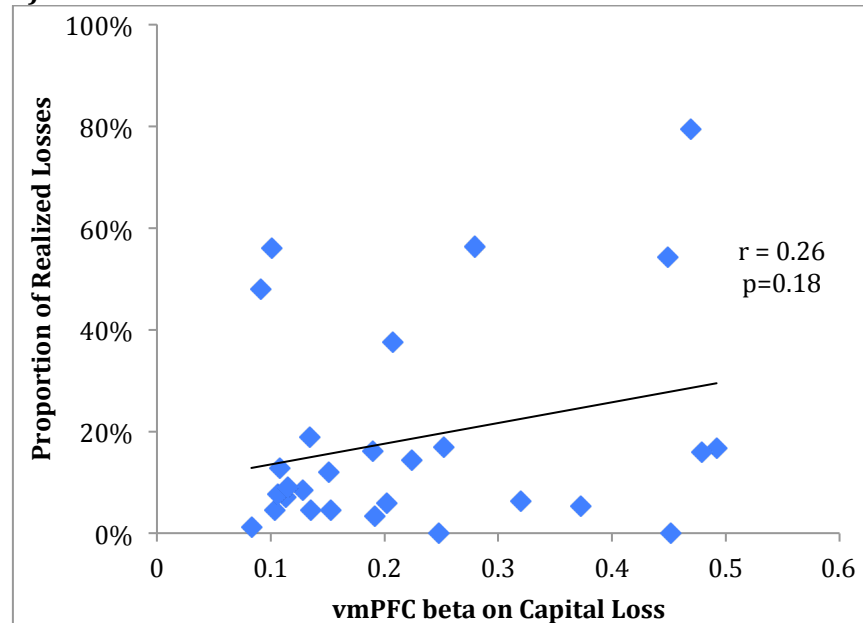
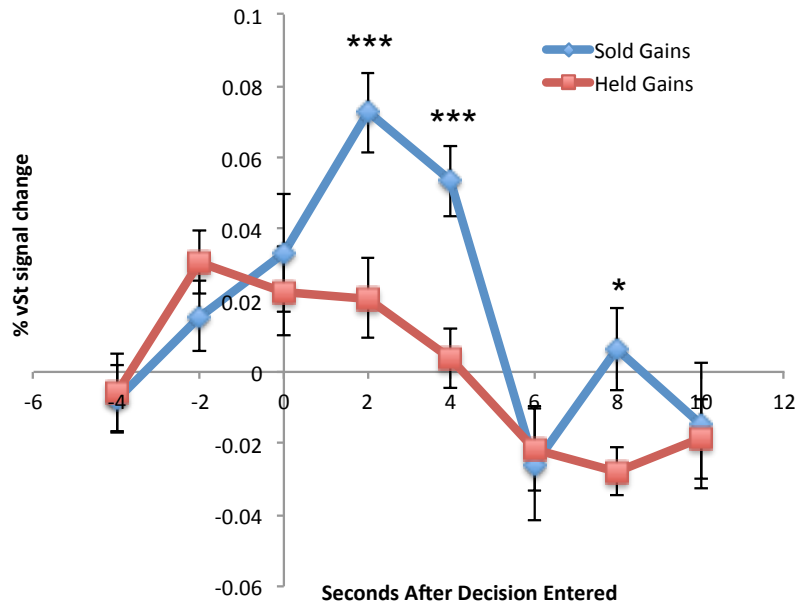
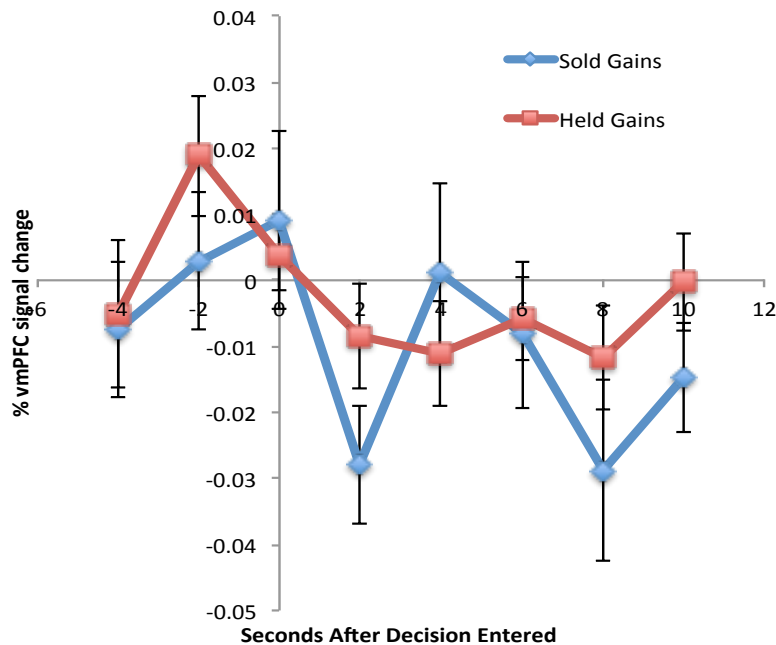


Figure 7. **Direct tests of the realization utility hypothesis.** Average activity in the vSt (Panel A) and vmPFC (Panel B) during trials when subjects were offered the opportunity to sell capital gains. The blue time series plots the average activity in trials where subjects realized capital gains, while the red time series plots the average activity in trials where subjects decided to hold capital gains. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$ (paired t-test). $t = 0$ corresponds to the instant at which the subject enters his trading decision on a hand-held device.

A)



B)



Appendix: Experimental Instructions

Buying your stock

In this experiment you will be given 350 experimental dollars to invest in three different stocks. Your job is to choose when to buy and sell each stock, so that you earn the most money by the end of the experiment. Throughout the experiment, you will see the price of each stock changing (more detail below), and you will use this information to decide when to buy and sell. When you sell a stock, you receive an amount of cash equal to the price of the stock. When you buy a stock, you receive one unit of the stock, but you must give up an amount of cash equal to the current price of the stock.

The three stocks you can buy or sell are simply called Stock A, Stock B, and Stock C. To begin the experiment you *MUST* buy all three stocks, where each stock costs \$100. Therefore, after you buy the three stocks, you will own one unit of each stock and have a total of \$50 remaining. For the remainder of the experiment, you are only allowed to hold a maximum of 1 unit of each stock, and you cannot hold negative units (no short selling.) However, you can carry a *negative cash balance* by buying a stock for more money than you have, but any negative cash balances will be deducted from your final earnings.

Structure of the market

In the experiment, you will see two types of screens, a *price update screen* and an *action screen*. In the price update screen, one stock will be randomly selected and you will be told if the selected stock price has gone up or down, and by how much. Note that you will only see an update for *one* stock at a time. You will not be asked to do anything during this screen, you will simply see information about the change in price.

Following the price update screen, another stock will be randomly chosen (it may be the same one you just saw) and you will be asked to take an action. If you currently hold a unit of the stock, you will be asked if you would like to sell the stock at the current price. If you do not currently own a unit of the stock, you will be asked if you would like to buy a unit at the current price.

The experiment will start out with 9 consecutive price update screens, and then you will have the opportunity to buy or sell after each subsequent price update screen.

How the stock prices change

Each stock changes price according to the exact same rule. Each stock is either in a good state or in a bad state. In the good state, the stock goes up with 55% chance, and it goes down with 45% chance. In the bad state, the stock goes down with 55% chance and it goes up with 45% chance.

Once it is determined whether the price will go up or down, the *size* of the change is always random, and will either be \$5, \$10, or \$15. For example, in the bad state, the stock will go down with 55% chance, and the amount it goes down by is \$5, \$10, or \$15 with equal chance. Similarly, the good stock will go up with 55% chance, and the amount it goes up by will either be \$5, \$10, or \$15.

The stocks will all randomly start in either the good state or bad state, and after each price update, there is a 20% chance the stock switches state.

Stock price changes

	<i>Good state</i>	<i>Bad state</i>
+	55%	45%
-	45%	55%

State changes

	<i>Good state today</i>	<i>Bad state today</i>
<i>Good state tomorrow</i>	80%	20%
<i>Bad state tomorrow</i>	20%	80%

Earnings and payout

You will play this market game **TWO SEPARATE TIMES** in the scanner. Each game will last approximately 15 minutes, and each game is independent from the previous one. This means when you start the second game, you will have to buy the three stocks at \$100 again, and the stocks will start randomly in each state again.

Your earnings at the end of the experiment will be equal to the amount of cash you accrued over the two scanning sessions from buying and selling stocks, plus the current price of any stocks that you own.

$$\text{Earnings} = \text{cash} + \text{price } A * (\text{Hold } A) + \text{Price } B * (\text{Hold } B) + \text{Price } C * (\text{Hold } C)$$

Finally, your earnings will be converted using an exchange rate of 12:1. That means we divide your earnings by 12, and pay you this amount plus the \$15 show up fee.

Button presses

During the Action screens, you will either be given the option to “Buy?” or “Sell?” depending on whether you hold the stock or not. The LEFT (blue) button indicates “YES”. And the RIGHT (yellow) button indicates “NO.” You have three seconds to enter your response, otherwise the computer will randomly select a response for you.