What Drives Peer Effects in Financial Decision-Making? Neural and Behavioral Evidence

CARY FRYDMAN *

February 2015

ABSTRACT

I use neural data collected from an experimental asset market to test the underlying mechanisms that generate peer effects. In a sample of randomly assigned subjects who are given identical information, I find strong causal peer effects in individual investment decisions. I then use the neural data to construct novel empirical tests that can separately identify a social learning mechanism and a relative wealth mechanism. The observed neural activity indicates that both mechanisms contribute to peer effects. I then construct a novel variable using individual measures of neural activity to explain cross-sectional variation in the size of the peer effects.

*contact: cfrydman@marshall.usc.edu, Marshall School of Business, University of Southern California.
Financial support from a Zumberge Individual Award Grant is gratefully acknowledged. I thank Kenneth Ahern, Tom Chang, Chad Kendall, Shimon Kogan, Camelia Kuhnen, Lawrence Jin, Matthew Shum, Alec Smith, David Solomon and seminar participants from Claremont, Cornell, USC, and the 2015 AEA meetings for helpful comments. I am also extremely grateful to Terry Lohrenz and Read Montague for sharing their neuroimaging data set.

Most models of trading in financial markets do not allow investors to socially interact with each other. Indeed, the primary mechanism in most models through which one investor can affect the beliefs and decisions of another investor is through the market price. In reality, however, individuals observe each others' behavior directly or learn about each others' decisions and beliefs through conversation (Hirshleifer and Teoh (2009)). Over the last fifteen years, empiricists have shown that these social interactions can have a significant impact on financial decision-making in a wide variety of contexts. For example, peers can affect stock market participation (Hong et al. (2004); Brown et al. (2008)), retirement saving decisions (Duflo and Saez (2003); Beshears et al. (2014)), mutual fund manager stock selection (Hong et al. (2005); Pool et al. (2014)), and individual investor trading decisions (Ivković and Weisbenner (2007); Bursztyn et al. (2014)).

Despite the growing body of evidence that documents peer effects in financial decision-making, there has been less success in understanding the *mechanism* that generates peer effects. Broadly speaking, there are two competing channels through which peers can affect financial decisions. First, an agent can learn about the fundamental value of an asset from a peer's decision (the information channel). Second, a peer's decision may directly enter another agent's utility function because of relative wealth concerns or a direct preference for conformity (the preference channel). In addition to guiding the development of newer models with socially interacting investors (Hirshleifer (2014)), identifying the different channels of peer effects is important because it can inform policy decisions on how to effectively curb herding behavior. However, distinguishing between the two channels is challenging because in many environments, an information-based mechanism makes similar predictions to a preference-based mechanism[1].

[1] For example, suppose an investor observes his neighbor experience a large positive return in the stock market, and subsequently increases his investment in the stock market (Kaustia and Knüpfer (2012)). This can occur through an information channel if the investor learns about fundamentals from his neighbor. However, it could also occur through a preference channel if the investor increases his investment in order to "Keep up with the Joneses."

In this paper, I take up the challenge of identifying the mechanisms that generate peer effects using a newly available type of data: neural activity measured with functional magnetic resonance imaging (fMRI). The neural data is valuable in addressing this challenge for two main reasons. First, while the two mechanisms often make similar predictions about trading behavior, I show that in the current setting, they make distinct predictions about the neural data generated during social interaction. Neural activity can therefore be viewed as an additional data source, one that enables me test competing theories of investor behavior that would be difficult to test using trading data alone. Second, while there is substantial heterogeneity in the magnitude of peer effects across the population, there has been little work on explaining the source of this heterogeneity (although, see Bursztyn et al. (2014)). I show that individual measures of neural activity can explain a significant portion of this cross-sectional variation. This result illustrates how empirical methods from neuroscience can be used to construct novel variables that explain variation in financial decision-making across individuals.

The data analyzed here are collected in a highly controlled laboratory experiment where subjects are randomly assigned into peer groups. This setup is attractive because it allows me to entirely sidestep the well-known identification problem where correlated behavior among two agents may be driven by (i) selection into peer groups or by (ii) common shocks within peer groups (Manski (1993)). Following the recent work on peer effects in finance by Lerner and Malmendier (2013), Shue (2013), and Ahern et al. (2014), I use the random assignment condition to solve the selection problem. Furthermore, and in contrast to previous work, I am able to perfectly observe all information that subjects see and all information that is transmitted between subjects. This allows me to control for common shocks to a peer group, and thus, I can rule out the two-part identification problem by design.

The methodology used here to test competing mechanisms of peer effects relies heavily on neural data; the details of both the neural predictions and the neural data are described later in the paper, but I summarize the main idea here. All of the neural tests rely on the theory of

prediction error, which is a signal that measures the change in expected net present value of lifetime utility generated by new information. Critically, a large body of evidence from cognitive neuroscience has shown that prediction errors can be measured in a specific area of the brain called the ventral Striatum (vSt)[2]. For economists, the potential value of this measurement technique should be clear as it implies that one can infer which factors – in addition to the standard consumption factor – affect a subject's utility[3].

Using this theory from neuroscience, the logic behind empirically testing the economic mechanisms is as follows. For each competing mechanism, I derive the change in discounted utility that is generated when a subject receives information about his peer's decision. For the social learning mechanism, denote this quantity by the function *f(x),* and for the social preference mechanism, denote this quantity by the function *g(x).* The vector *x* contains the history of investment allocations and the current wealth levels of each subject. The functions *f(x)* and *g(x)* therefore represent the prediction errors that are generated under the social learning and social preference channels, respectively. By observing neural activity in the vSt – the precise area in which prediction errors are encoded – I can test whether this activity correlates with the hypothesized prediction errors. A significant correlation between vSt activity and *f(x)* suggests that the social learning channel is active, while a significant correlation between vSt activity and *g(x)* suggests that the social preference channel is active.

Although there are a variety of mechanisms within the class of social preferences, the above methodology is general enough that it can infer specific mechanisms within the broader class. In particular, different types of social preference mechanisms will generate different functional forms for *g(x),* allowing me to infer specific mechanisms. In this paper, I mainly focus on a relative wealth mechanism because of its prevalence in the finance literature, (Abel (1990);

[2] See, for example, Schultz et al. (1997), Pessiglione et al. (2006), and Caplin et al. (2010).
[3] Indeed, a series of recent papers in neuroeconomics has highlighted the vast potential that this technique holds for answering fundamental questions about choice theory and belief updating (Caplin and Dean (2008); Caplin et al. (2010); Rutledge et al. (2010)).

Gali (1994)). However, later in the paper I also show how this same method can test for a different preference-based mechanism, one driven by a taste for conformity.

Beyond the standard difficulties in testing channels of peer effects, one additional complication with testing for relative wealth concerns is that changes in absolute wealth are often correlated with changes in relative wealth. However, a key feature of the experimental design is that information about the market return is revealed before information about the peer decision is revealed. This staggered information arrival allows me to construct neural tests that can separately identify whether a change in utility is driven by absolute changes in wealth or relative changes in wealth.

The main results of the paper can be summarized as follows. First, neural activity collected during social interaction provides novel evidence that a relative wealth mechanism and a social learning mechanism are both significant in contributing to the observed peer effects. Second, the random assignment of peers coupled with the high level of experimental control in a lab environment enable me to precisely estimate causal peer effects at the individual subject level; this, in turn, allows me to characterize the distribution of peer effects across the subject pool. Finally, I show that cross-sectional variation in peer effects is significantly explained by individual measures of neural activity generated during social interaction. On a methodological level, these results also provide an example of using a novel empirical framework, one that can be used in future work to test competing mechanisms of other types of economic behavior besides peer effects.

My results mainly contribute to the literature on peer effects and to the growing literature on neurofinance. As recent empirical work has achieved growing success in cleanly identifying causal peer effects, the natural next step is to ask, what drives these peer effects? There has been little work on this question, with the exception of a recent paper that uses a field study in Brazil (Burszstyn et al. (2014)). These authors use a clever experimental design to separate an investor's revealed preference for a risky asset from the ownership of the risky asset. They can then infer

whether investors are purchasing an asset due to social learning (information) or social utility (preferences). Burszstyn et al. (2014) find that both mechanisms explain a portion of the observed peer effects, which is consistent with the neural results reported here.

Although the current study is also aimed at teasing apart the different mechanisms, the neural data used here offers the extra ability to directly test specific mechanisms within the class of social preferences. While Burzstyn et al. (2014) focus on a mechanism where an investor's utility for owning an asset is increasing in his peer's ownership, I show that this may not always be the case. In particular, I provide evidence that when an asset delivers a negative return, a subject derives a negative utility shock if he invested more in the asset than his peer did. Another difference between the two studies is that subjects in the current experiment are randomly assigned to peer groups and never meet each other face-to-face, whereas investors in Burszstyn et al. (2014) form their peer groups endogenously. My finding that there is a significant correlation in investment behavior between *anonymous* randomly assigned peers suggests that the two peer effects mechanisms are strong enough to operate even in the absence of prior social interaction.

This paper also contributes to the young but growing field of neurofinance. Many of the early contributions to this field provide important evidence showing the neural and physiological correlates of financial risk-taking (Lo and Repin (2002); Kuhnen and Knutson (2005); Preuschoff et al. (2006)). Subsequent studies build on this earlier work and demonstrate how exogenous emotional cues causally impact financial risk-taking (Knutson et al. (2008); Sokol-Hessner et al. (2009)).

While recent work in this area is starting to investigate a wider array of topics such as learning and the formation of asset price bubbles,[4] this paper contributes to neurofinance by providing a basic methodology that can be used for testing competing mechanisms of economic behavior. In particular, I provide an example of how using the theory of prediction errors from

4 See Bossaerts and Le Nestour (2014) and Kuhnen (2014) for work on learning in financial markets, while De Martino et al. (2013) and Smith et al. (2014) explore the neural mechanisms of experimental asset price bubbles.

computational neuroscience can provide a new set of testable predictions for models of financial decision-making. As such, this paper is related to two recent studies that use neural data to test theories of investor behavior. One uses neural activity generated from selling stocks to test the "realization utility" theory of trading (Frydman et al. (2014)); the other uses vSt activity generated upon viewing stock returns to test whether regret theory can explain stock purchasing behavior (Frydman and Camerer (2014)).

I. Experimental Design and Predictions

In this section, I first describe the experimental stock market that was used to generate both the behavioral and neural data. I then outline the neural predictions that are generated under the competing mechanisms that can give rise to peer effects.

A. Design

Both the design and all the data analyzed here are taken from the experiment conducted in Lohrenz et al. (2013), and I now describe their experimental setting in detail. 48 subjects are endowed with $100 of experimental cash and given the opportunity to invest this wealth in two separate assets over the course of two hundred trials. One asset is risk-free and pays a zero interest rate, and the other asset is risky and generats a random return R_t in each period $t=1,2,...,200$. The sequence of risky asset returns $\{R_t\}$ is taken from historical markets, and subjects are not told anything about the process $\{R_t\}$, except that it is taken from real historical markets. Subjects are therefore faced with a large amount of parameter uncertainty upon entering the experiment. At the beginning of each period t, subject i allocates a fraction of his wealth, $x_{i,t}$, to the risky asset. The remaining fraction of wealth, $(1 - x_{i,t})$, is invested in the risk-free asset that earns a zero interest rate.

A sequence $\{R_t\}$ of length twenty is referred to as a "market", and each subject participates in ten separate markets throughout the course of the experiment. Subjects are instructed that each market is mutually independent of the other nine markets and are given a short break in between each market; at the beginning of a new market, a subject's wealth is carried over from the end of the previous market. Upon entering the experiment, half of the 48 subjects are randomly assigned to a *social treatment,* and the remaining 24 subjects are assigned to a *control condition.* Subjects within the *social treatment* are then randomly assigned into twelve pairs.

Each of the two hundred trials consists of three different screens (Figure 1). First, a subject sees an "allocation" screen at which time he is instructed to enter his investment allocation $x_{i,t}$. Second, a "market" screen displays both the realized return of the risky asset and the subject's updated portfolio value as a function of his investment allocation. Finally, if the subject is in the *social treatment,* a "peer decision" screen reveals the investment allocation of subject *i's* peer. If instead the subject is in the *control condition,* the "peer decision" screen reveals a randomly drawn investment allocation that is uniformly drawn from [0,1]. Subjects in the *control condition* are explicitly told that the investment allocation they observe on the "peer decision" screen is indeed uniformly drawn from [0, 1]. At the end of the experiment, a subject's final portfolio is converted from experimental currency to actual US dollars using a 5:1 exchange rate. In addition to the earnings from the experiment, subjects are paid a fixed "show-up" fee of $20.

B. Behavioral Predictions

As described in the introduction, there are well-known difficulties with identifying causal peer effects in economic behavior (Manski (1993)). Correlated behavior among two agents can potentially be driven by common shocks to a peer group or by endogenous selection into a peer group. For example, two neighbors may hold similar portfolios because they invest based on the

same information from a local newspaper. Alternatively, two investors may select into the same neighborhood based on similar levels of risk aversion, and this selection may account for the observed similarity in portfolios. In the current experiment, random assignment of peers combined with the ability to observe all information shown to each subject, is sufficient to rule out the identification problems by design.

In order to test for the presence of peer effects, it is useful to provide a precise definition of what constitutes a peer effect in the current experimental setting. A peer effect arises in the social condition when $x_{i,t}$ is dependent on the history of peer investment allocations $\{x_{j,u}\}_{u=1,\ldots,t-1}$ controlling for any function of past returns $\{R_u\}_{u=1,\ldots,t-1}$. For simplicity, the empirical tests I construct will invoke a stronger definition, where a peer effect requires $x_{i,t}$ to be dependent on the most recent peer investment allocation, $x_{j,t-1}$. It is necessary to control for all functions of past returns because it is possible that investment decisions are correlated within pairs due to these common shocks.

While I cannot explicitly control for every function of past returns, I can exploit the fact that, within each market, subject *k* in the *control condition* sees the same sequence of returns as subject *i* and *j* in the *social treatment* (where *i* and *j* belong to the same pair). Therefore, in the absence of peer effects I expect that on average, $corr(x_{i,t}, x_{j,t-1}) = corr(x_{i,t}, x_{k,t-1})$. In other words, without peer effects, there should be no difference between the within-pair and across-pair correlations in investment decisions. In contrast, if there is a causal effect of *j's* investment allocation in period *t-1* on *i's* investment allocation in period *t,* this can be detected by a difference between the within-pair and across-pair correlations. This leads to the first prediction:

Prediction 1 (Behavioral): Let *i* and *j* be randomly assigned peers in the *social treatment*, and let *k* be any subject in the *control condition.* If there are no peer effects, then the within-pair correlation should on average, equal the across-pair correlation: $corr(x_{i,t}, x_{j,t-1}) = corr(x_{i,t}, x_{k,t-1})$. In contrast, if there are peer effects, then on average, $corr(x_{i,t}, x_{j,t-1}) \neq corr(x_{i,t}, x_{k,t-1})$.

C. Neural Predictions

While Prediction 1 is concerned with identifying causal peer effects, this section develops the predictions regarding the different mechanisms that can generate peer effects. In particular, I exploit the availability of neural data at the peer decision screen in order to generate testable predictions of competing theories.

Broadly speaking, there are two main channels through which peer effects can arise. First, peer decisions may be correlated if *i* learns about an asset's return distribution from *j's* decision. This is often referred to as observational learning (Bikhchandani et al. (1992)) and acts through an information channel. Second, peer effects may arise if *i*'s utility depends directly on *j*'s decision. In what follows, I focus on a specific mechanism within the class of social utility mechanisms where *i* has utility over relative wealth compared to *j.* Later on, I provide a neural test of a different social utility mechanism, one based on a taste for conformity.

The focus of the neural predictions is on neural activity generated at the moment when a peer's investment allocation is revealed. Recall that when this "peer decision" screen appears, a subject observes his peer's investment allocation but is not required to make any active decisions. Hence, this screen is exclusively involved with information processing, as opposed to decision deliberation. This is useful because a large literature in decision neuroscience has been devoted to precisely characterizing the neural activity that is generated when new information -- in this case, a peer's investment decision -- is revealed. Specifically, I focus on the concept of a *prediction error*, which I briefly review now.

A prediction error is a signal that the brain computes in response to new information. This signal can be thought of as measuring the change in the expected net present value of utility generated by the news, taking into account all sources of utility. While this terminology may be unfamiliar to most financial economists, the general concept of a prediction error should be familiar as it is closely related to basic properties of asset pricing. To see this, consider a simple

model where an asset's price is equal to the sum of its discounted expected cash flows. Prices should then fluctuate only when unexpected news is revealed about future cash flows, and this price fluctuation should occur at the moment when the unexpected news is revealed. Furthermore, the sign and size of the price change should reflect the amount by which the cash flows are expected to increase or decrease. Similarly, a prediction error signal embodies these same two core properties that a price change exhibits in an efficient market: (i) it is different from zero only at the time when unexpected news is revealed and (ii) it carries information about the signed change in expected utility generated from the unexpected news.

Critically, a large body of evidence in decision neuroscience shows that the prediction error signal can be accurately measured in a specific area of the brain, the vSt (Schultz et al. (1997); Pessiglione et al. (2006); Hare et al. (2008); Lin et al. (2012)). This is useful because it allows me to empirically measure the prediction error signal generated in response to news, and therefore, infer which factors cause a change in discounted expected utility. To fix ideas, for a given mechanism ***M*** I denote the change in expected discounted utility that is generated by news as the function *m(x);* this function represents the time series of prediction errors that is theoretically predicted by mechanism ***M.*** I can then test whether the time series of vSt activity significantly correlates with *m(x)*; if it does, this suggests that the mechanism used to generate the prediction errors – and summarized by the *m(x)* function -- is a significant contributor to economic behavior. I now develop the specific prediction error functions that are used to test between the competing mechanisms of peer effects.

First, I construct the prediction error function, denoted by *f(x),* that is generated by the social learning mechanism. While all subjects are given the same information regarding the markets at the start of the experiment, one cannot rule out *a priori* the possibility that *i* may perceive his peer, *j,* to be more financially sophisticated than he is. For example, if *i* has never taken a finance class or invested in the stock market, he may believe that on average, others in the population will have more valuable information than he does. Hence, it is possible that even

though *i* and *j* each see the same sequence of returns, *i* may believe that *j*'s decision carries information about the underlying distribution of returns.

In this case, *i* should update upwards (downwards) his belief about the value of the risky asset after observing *j* increase (decrease) the fraction of his wealth allocated to the risky asset. All things equal, when subject *i* updates his beliefs upward about the value of the asset, this news should increase his discounted expected utility as it implies that the risky asset is a better investment opportunity than previously expected. Therefore, the change in *j*'s risky asset allocation, $(x_{j,t} - x_{j,t-1})$, is a prediction error, and this quantity should correlate with vSt activity at the moment when the peer decision screen is revealed[5].

Prediction 2 (Neural Learning Mechanism): If peer effects are driven by a social learning mechanism, then at the time when a peer decision is revealed, vSt activity should positively correlate with the prediction error function defined by: $f(x) = (x_{j,t} - x_{j,t-1}) - (x_{i,t} - x_{i,t-1})$.

Next, I turn to Prediction 3, one that characterizes the neural activity that should be observed if peer effects are generated through a social preference channel. As before, the goal here is to construct a function that summarizes the prediction errors generated by the social preference mechanism. While there a variety of different mechanisms within the class of social preferences, I focus on "keeping up with the Jones'" preferences, where subjects have utility over relative wealth (Abel (1990); Gali (1994)). As such, this particular mechanism will generate a

[5] Before stating the precise definition of this neural prediction, a small adjustment is needed because the market return in each trial is revealed before the peer decision is revealed. To understand why this adjustment is necessary, consider the peer decision revealed in trial *t*. When subject *i* observes $x_{j,t}$, he is observing how *j* changed his investment allocation as a function of the market return on trial *t-1*. However, subject *i* has also updated his beliefs about the risky asset as a function of the market return on trial *t-1*, and this should be captured by the change in his investment allocation $(x_{i,t} - x_{i,t-1})$. Therefore, to control for the updating that subject *i* does through private learning, I subtract this quantity from the change in *j's* risky asset allocation. The prediction error attributed to social learning by subject *i* on trial *t* is therefore equal to: $(x_{j,t} - x_{j,t-1}) - (x_{i,t} - x_{i,t-1})$. Note that this specification assumes that subject *i* is naïve in the sense that he believes his peer will change an investment allocation only as a function of market returns, and hence not as a function of *i's* decision (Eyster and Rabin (2010)).

specific functional form which I denote by $g^{Jones}(x)$, that will be different from the functional forms generated by other mechanisms within the class of social preferences. It is worth highlighting that there is no incentive in the experimental design that induces subjects to have a concern for relative wealth; therefore, the test of this mechanism relies on exogenous preferences over relative wealth[6].

Under this mechanism, subjects derive utility from the difference between their peer's final portfolio value and their own final portfolio value[7]. This implies that the change in *j's* portfolio value on each trial contains news about *i's* expected discounted utility. Therefore, the revelation of the peer decision (along with the risky asset return) should generate a prediction error for subject *i.* Recall that subjects are not explicitly shown their peer's portfolio value on each peer decision screen, and so it may be difficult for them to compute the actual dollar value change in their peer's portfolio each period. However, subjects are shown the fraction of wealth their peer allocated to the risky asset, $x_{j,t}$, and the risky asset return R_t on each trial. It is then reasonable to expect that subjects can compute both the percentage change in their peer's portfolio value, and the percentage change in their own portfolio value on each trial.

I therefore proxy the dollar change in portfolio values with the percentage change in portfolio values. The prediction error generated under the relative wealth mechanism should then be equal to the difference between the percentage change in *i's* portfolio value and the percentage change in *j's* portfolio value. That is, the prediction error function generated by a relative wealth concern mechanism is defined by, $g^{Jones}(x) = R_t \times (x_{i,t} - x_{j,t})$.

6 This is in contrast to relative wealth concerns that can arise endogenously when one consumes scarce resources in the economy (DeMarzo et al. (2008)).

7 This, of course, does not preclude subjects from having utility directly over their own final wealth. For example, a subject could have utility of the form $u(m, y) = m + g(m-y)$, where m is a subject's own wealth, y is the amount of peer wealth, an g is an increasing function.

Prediction 3 (Neural Preference Mechanism): If peer effects are driven by a relative wealth mechanism, then at the time when a peer decision is revealed, vSt activity should positively correlate with the prediction error function defined by: $g^{Jones}(x) = R_t \times (x_{i,t} - x_{j,t})$.

The previous two predictions indicate that if a causal peer effect is observed, then neural activity in the vSt should correlate with (i) a preference-based prediction error, (ii) an information-based prediction error, or (iii) both types of prediction errors. Moreover, the strength with which the vSt encodes these prediction errors should correlate with the size of the observed peer effects. In other words, subjects whose vSt activity more strongly correlates with the prediction errors defined by $f(x)$ and $g^{Jones}(x)$ should be more affected by their peer's investment allocations. This leads to the final prediction, which provides a link between the individual measures of neural activity and the individual measures of peer effects.

Prediction 4 (Neural Cross-Sectional Variation): The degree to which vSt activity correlates the predictions errors $f(x)$ and $g^{Jones}(x)$, should itself be correlated, across subjects, with the size of the peer effect.

II. fMRI Data Collection and Analysis

In this section, I describe how the fMRI measures of neural activity are collected and analyzed. The goal in this section, which is taken primarily from Frydman et al. (2014), is to provide a brief primer on the basics of fMRI data analysis, so that the basic econometrics are approachable to those not already familiar with the fMRI literature. For a more detailed discussion, see Huettel, Song, and McCarthy (2004), Ashby (2011), and Poldrack, Mumford, and Nichols (2011).

A. fMRI Data Collection and Measurement

Measures of neural activity are collected 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 the local inflows of oxygenated hemoglobin and outflows of de-oxygenated hemoglobin that occur when neurons fire. In particular, fMRI provides measures of the BOLD response in small "neighborhoods" of brain tissue called *voxels*, and is thought to measure the sum of the total amount of neuronal firing into that voxel and the total amount of neuronal firing within the voxel.

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 response takes place quickly, BOLD measurements are affected for up to 24 seconds thereafter. Figure 2A provides a more detailed illustration of the nature of the BOLD response. The top panel depicts the path of the BOLD signal in response to one (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 time elapsed 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, there is a standard way of dealing with the complication described in the previous paragraph. In particular, the BOLD response has been shown to combine linearly across multiple sources of neural activity (Boynton et al. (1996)). This property, along with knowledge of the specific functional form of the HRF, allows us to construct a mapping from predicted neural activity to predicted BOLD responses. Specifically, if the predicted 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

$$b(t)=\int_0^{\infty} h(u)a(t-u)\,du, \tag{1}$$

which is the convolution between the HRF and the neural inputs. This integral has a straightforward interpretation: it is a lagged sum of all the BOLD responses triggered by previous neural activity. The properties of the BOLD response are illustrated in Fig. 2B, which depicts a hypothetical path of neural activity (solid line), together with the associated BOLD response (dashed line).

During the experiment, two types of MRI data are acquired in a 3.0 Siemens Tesla Trio MRI. First, BOLD-fMRI data is acquired while the subjects perform the experimental task. The voxel size used is 3.4 mm x 3.4 mm x 4 mm, and this data is collected for the entire brain every 2 seconds. Additionally, high-resolution anatomical scans are acquired and are used mainly for realigning the brains across subjects and for localizing the brain activity identified by the analyses.

B. fMRI Main Data Analyses

The key goals of the analysis are to test if neural activity in the region of the vSt that has been repeatedly shown to encode prediction errors is consistent with neural predictions 2, 3 and 4. To do this, I run statistical tests to see if there are areas within this region 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 (prediction errors generated by social learning or relative wealth mechanisms). This is complicated by the fact that, since every voxel contains thousands of neurons, the BOLD responses in a voxel can be driven by multiple signals. Fortunately, the linear properties of the BOLD signal allow the neural signals of interest to be identified using standard linear regression methods.

The general statistical 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 I look for (e.g., a prediction error at the revelation of a peer decision). 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 in the BOLD signal, but are not explicitly used in any of my tests. The control variables are not convolved with the HRF because, while they affect the measured BOLD responses, they do not reflect neural activity which triggers a hemodynamic response.

The linearity of the BOLD signal implies that the level of BOLD activity $b^v(t)$ in any voxel v at time t should be given by

$$b^v(t) = \text{constant} + \sum_i \beta_i^v S_i(t) + \sum_j \alpha_j^v c_j(t) + \varepsilon(t), \tag{2}$$

where $\varepsilon(t)$ denotes AR(1) noise. This model is estimated independently in each of the voxels that fall within vSt. My 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 this one. First, I am interested in testing hypotheses about the distribution of the signal coefficients in the population of subjects, not hypotheses about individual subject coefficients. This would normally require estimating a mixed effects version of the linear model specified above, which, given the size of a typical fMRI dataset, would be computationally intensive. Fortunately, there is a 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 I follow here.

Second, I conduct tests in an area of the vSt that, in prior work, has been linked to the computation of prediction errors. Specifically, I construct two 8 mm radius spheres (a total of 56 voxels) around the coordinates (MNI-space, $[x = -10, y = 12, z = -8]$, $[x = 10, y = 12, z = -8]$) that are found to exhibit peak correlation with prediction errors in Pessiglione et al. (2006). Because my hypothesis tests are therefore carried out in each of the 56 voxels in the relevant vSt region of interest, there is a concern about false-positives. To address this problem, I correct for multiple comparisons within the relevant region of interest, a procedure known in the fMRI literature as a small volume correction (SVC). I report results as significant if they pass SVC correction at a level of $p<0.05$[8].

III. Results

A. Test of Behavioral Prediction 1

I begin the test of Prediction 1 by computing, for each subject *i* in the *social treatment,* the correlation between his risky asset allocation in period *t,* $x_{i,t}$, and his peer's risky asset allocation in period *t-1,* $x_{j,t-1}$. The average value of $corr(x_{i,t}, x_{j,t-1})$ across all twenty-four subjects in the *social treatment* is 0.137 (standard error: 0.026). Next, I compute the correlations generated by the risky asset allocations of subject *i* in the *social treatment* and subject *k* in the *control condition.* Specifically, I fix subject *i* in the *social treatment,* and then compute $corr(x_{i,t}, x_{k,t-1})$, for each of the *k=1,…,24* subjects in the *control condition.* I then average these twenty-four

[8] Specifically, I report results as significant if voxels pass SVC with a family-wise error rate of less than 0.05 and if they pass $p<0.005$ uncorrected with a 10-voxel extent threshold.

correlations, and take this to be the across-pair correlation measure for subject *i*. Finally, I take the average of this across all twenty-four subjects in the *social treatment,* which equals 0.063 (standard error: 0.013).

I can reject the null hypothesis that the within-pair and across-pair correlations are equal with a *t*-statistic of 2.93. Figure 3 shows that for sixteen of the twenty-four subjects in the *social treatment,* the within-pair correlation is greater than the across-pair correlation[9]. Because the experimental design ensures that subjects in the *control condition* and *social treatment* observe the same history of returns, the result that the within-pair correlation is significantly greater than the across-pair correlation provides evidence that peer decisions causally affect investment allocations.

To gain further insight into other factors that explain the time series variation in risky asset allocations, I impose additional structure on the model that subjects use to form their allocations. Because subjects are not told anything about the actual data generating process that governs the risky asset returns, they are faced with a learning problem. In such a setting, a natural hypothesis is that a subject computes the conditional Sharpe ratio after observing market returns and then uses this statistic to guide the investment decision.

To test this hypothesis, I estimate an OLS regression of subject *i's* current investment allocation on the conditional Sharpe ratio, his peer's previous investment allocation, and subject *i's* previous investment allocation:

$$x_{i,t} = \alpha + \gamma_i + \beta_1 x_{j,t-1} + \beta_2 sharpe_ratio_t + \beta_3 x_{i,t-1} + \varepsilon_{i,t} \qquad (3)$$

The model is estimated separately for the *social treatment* and for the *control condition,* and results are displayed in Table 1. The first column of Table 1 confirms the previous univariate correlation results, as *j's* previous investment allocation is a significant predictor of *i's* current investment allocation. Furthermore, column 1 and column 2 both show that the conditional

[9] Note also that the average *across-pair* correlation is significantly greater than zero (*t*-statistic of 4.54). This suggests that subjects exhibit correlated behavior, in part, due to exposure to common information.

Sharpe ratio is a significant predictor of risky asset allocations. Therefore, in the *social treatment,* variation in risky asset allocations can be decomposed into a social component (the peer investment decision) and a non-social component (the conditional Sharpe ratio)[10]. Interestingly, when pooling data across control and treatment groups (column 3), the interaction between the treatment and the conditional Sharpe ratio has a negative coefficient (significant at the 10% level), suggesting that when peer decisions are observable, subjects rely less on past returns.

While the regression results presented above are consistent with peer effects, one alternative theory that can potentially explain part of the observed behavior is anchoring (Tversky and Kahneman (1974)). Anchoring occurs when individuals rely on arbitrary initial values or starting points when computing decisions in complex environments. The subjects here may perceive the experiment as a complex environment, and peer decisions can provide anchors for subjects in the *social treatment* when they are choosing their risky asset allocations. In this case, one would expect to see the observed positive correlation between *j's* allocation on trial *t-1* and *i's* allocation on trial *t.*

However, I can test this alternative theory using the *control condition* data in Table 1. Recall that in the *control condition,* subjects do not have access to a peer's decision on each trial, but instead see a randomly drawn number over the interval [0,1] on the peer decision screen. Hence, if peer effects are driven by anchoring, I should still observe a correlation between the uniformly drawn number (the anchor) and the risky asset allocation. Instead, the coefficient on the peer investment is not significantly different from zero in the *control condition*, and furthermore, the coefficient on the interaction between the social condition and the peer investment (column 3) is significantly positive. Together, these results cast doubt on the anchoring hypothesis.

[10] The fact that the conditional Sharpe ratio is a significant predictor of investment decisions is consistent with the positive *across-pair* correlations we document in Figure 3. In other words, the conditional Sharpe ratio may be the channel through which two subjects exhibit correlated decisions, despite their inability to observe each other's decisions.

B. Neural Response to Market Returns

Before turning to the results on neural predictions 2 and 3, I first describe a preliminary result on neural activity at the time when market returns are revealed. This preliminary result will act to validate the methodology used in the next section to test the key neural predictions. Recall that in each trial, a subject observes the market return before observing his peer's investment allocation (Figure 1). If a subject observes a positive (negative) market return, then conditional on investing in the risky asset, he will experience an increase (decrease) in wealth. This change in wealth carries news about future consumption, and hence a positive (negative) market return should generate a positive (negative) prediction error in the vSt. Therefore, when the market return is revealed to a subject, vSt activity should positively correlate with the subject's change in wealth, given by $R_t x_{i,t}$.

To check whether vSt activity is consistent with this prediction, I estimate the following general linear model (GLM) of BOLD activity in every subject and voxel:

(4)

$$b^v(t) = \alpha + \beta_1^v I_{PD}(t) \times R_t\left(x_{i,t} - x_{j,t}\right) + \beta_2^v I_{PD}(t) \times \left[\left(x_{j,t} - x_{j,t-1}\right) - \left(x_{i,t} - x_{i,t-1}\right)\right] + \beta_3^v I_{MKT}(t) \times \left(R_t x_{i,t}\right) + \beta_4^v controls + \varepsilon(t)$$

This is the same model that will be used to perform tests of predictions 2 and 3, so it is useful to understand its components in detail. $b^v(t)$ denotes the BOLD signal at time t in voxel v. I_{PD} is an indicator function that equals one if, at time t, the peer decision screen is revealed; I_{MKT} is an indicator function that equals one if, at time t, the market screen is revealed; the first nonconstant regressor represents the relative change in wealth that is revealed at the peer decision screen; the second nonconstant regressor represents the social learning prediction error described in section I.C. The third nonconstant regressor, $I_{MKT}(t) \times (R_t x_{i,t})$, represents the absolute change in wealth

that the subject experiences during the "market screen" at time *t.* The "controls" vector includes the following variables: 1) an indicator function denoting the onset of an allocation screen, 2) an indicator function denoting the onset of a market screen, 3) an indicator function denoting the onset of a market screen interacted with the risky asset return, 4) an indicator function denoting the onset of a peer decision screen, 5) an indicator function denoting the onset of a new market and 6) 12 movement regressors that control for subject head movement inside the scanner.

Controls (1) - (5) are convolved with the HRF, whereas control (6) is not. I include these controls because the BOLD signal is affected for up to 24 seconds after the initial neural impulse generated by the onset of any single event. Therefore, even though I am concerned only with the neural activity generated at the moment when a peer decision or market return is revealed, the observed BOLD signal at this time is still affected by the onset of several preceding screens.

The next step in testing whether vSt activity correlates with changes in wealth is to perform inference about the extent to which the signal of interest is encoded in a given voxel. This is done by carrying out a one-sided *t*-test against zero of the average of the individually estimated coefficients. In other words, I compute the average β_3^v across subjects and perform hypothesis tests under the null that the coefficient is zero; finally, I correct for multiple comparisons within the pre-specified vSt region of interest.

Figure 4 shows the results of this hypothesis. Within the pre-specified 56-voxel vSt target region (those voxels colored in yellow and orange) associated with the computation of prediction errors in previous studies, I find a cluster of 15 voxels (those voxels colored in orange and red) where β_3^v, averaged across subjects, is significantly positive ($p<0.05$ SVC). Because an increase in wealth should unambiguously increase expected discounted utility, this result acts to validate the neural methodology that I use to test the key neural predictions in the next section.

C. Test of Neural Predictions 2 and 3

I now turn to predictions 2 and 3, which investigate the extent to which neural activity in the vSt reflects prediction errors that are generated under the two different peer effects mechanisms outlined in section I.C. These are the key predictions of the paper, because they allow me to test specific mechanisms of peer effects that are difficult to test using trading data alone. I proceed in testing predictions 2 and 3 by performing hypothesis tests about the estimated coefficients from equation (4). Figure 5A shows that within the pre-specified 56-voxel vSt target region (those voxels colored in yellow and orange) associated with the computation of prediction errors in previous studies, there is a cluster of 9 voxels where β_2^v, averaged across subjects, is significantly positive ($p<0.05$ SVC). This indicates that at the time when a peer decision screen is revealed, activity in the vSt positively correlates with the theoretical prediction error that is generated under the social learning mechanism; hence the observed vSt activity is consistent with prediction 2.

The test of Prediction 3, which investigates the relative wealth mechanism, proceeds in a similar manner. Figure 5B shows that within the pre-specified 56-voxel vSt target region (again, those voxels colored in yellow and orange), I find a cluster of 19 voxels where β_1^v, averaged across subjects, is significantly negative ($p<0.05$ SVC). This indicates that at the time when a peer decision screen is revealed, activity in subject *i's* vSt significantly correlates with the change in his relative wealth.

One concern with the previous result is that changes in relative wealth can be highly correlated with changes in absolute wealth, especially if a peer invests only a small amount of wealth in the risky asset. Therefore, a prediction error that is consistent with a relative wealth mechanism may instead be driven by a change in absolute wealth. However, there is a key aspect of the experimental design that allows me to rule out this interpretation. Recall that a prediction error is generated only in response to unexpected information, and therefore, it will be equal to zero in response to information that has already been revealed in the past. In the experiment, a subject observes the market return before observing his peer's decision (Figure 1); it follows that

any change in expected discounted utility that is driven by a change in absolute wealth must be reflected in the prediction error at the time the market return is revealed. This is precisely the signal that is displayed in Figure 4. Therefore, the prediction error generated at the peer decision screen cannot be driven by changes in absolute wealth, because this information was previously revealed.

Taken together, the results from this section provide novel support in favor of both prediction 2 and prediction 3. It is important to point out that each mechanism explains a unique portion of the variation in vSt activity at the peer decision screen: the two key variables that parameterize the prediction errors of each mechanism are entered into the same general linear model, and are allowed to compete for explained variance. Therefore, these results suggest that peer effects are simultaneously driven through a preference channel and an information channel.

D. Test of Neural Prediction 4

If the observed peer effects are driven by a combination of the preference and information channels, then the neural signals identified in the previous section should explain a portion of the cross-sectional variation in these peer effects. Specifically, a subject whose vSt is particularly sensitive to the prediction error generated by his peer's investment should also exhibit a particularly strong peer effect.

To implement this test I first define a composite measure of a subject's vSt sensitivity to the information-based and preference-based prediction errors. For each subject, I compute (i) a neural measure of the degree to which the information-based prediction error, *f(x),* is represented in the vSt and (ii) a neural measure of the degree to which the relative wealth prediction error, $g^{Jones}(x)$, is represented in the vSt. These subject-specific neural measures are given by the average value, across voxels in the vSt target region, of β_1^v and β_2^v from equation (4); for subject *i,* these average beta values are denoted by $\beta_{i,1}^{avg}$ and $\beta_{i,2}^{avg}$. I then demean and normalize each of

these two values and sum them to attain a composite measure for each subject: $\beta_i^{composite} = zscore\left(\beta_{i,1}^{avg}\right) + zscore\left(\beta i,_2^{avg}\right)$. This composite measure therefore represents the strength with which the vSt responds to information contained in a peer's investment decision.

Consistent with prediction 4, figure 6 shows that there is a positive correlation, across subjects, between the neural composite measure and the size of the peer effect (*r=0.40, t*-statistic: *2.06*). The figure provides a scatterplot of the composite measure against the within pair correlation, $corr(x_{i,t}, x_{j,t-1})$, where each data point represents a subject in the *social treatment.* In summary, this result shows that there is significant heterogeneity in the size of peer effects, and a significant portion of this heterogeneity can be explained by individual measures of neural activity.

E. Neural Test of a Preference for Conformity

As mentioned in the introduction, there is a large class of social preference mechanisms that can generate peer effects. So far I have focused on a single mechanism based on a concern for relative wealth, but I now examine an alternative explanation based on a direct preference for conformity (Asch (1951)). Following the methodology for testing prediction 2 and 3, I derive the prediction error function that is generated under the conformity mechanism, which I denote by $g^{CONFORMITY}(x)$.

If subject *i* is influenced by subject *j* because of a direct preference for conformity, then on each trial, *i* should derive positive utility from observing that *j* made a similar investment allocation. In other words, at the time a peer decision is revealed, a direct taste for conformity should generate utility that is decreasing in the distance between subject *i's* allocation and subject *j's* allocation. As such, when the peer decision is revealed, the vSt activity of subject *i* should negatively correlate with $|x_{i,t} - x_{j,t}|$. This means that the prediction error function generated by the conformity mechanism is given by, $g^{CONFORMITY}(x) = -|x_{i,t} - x_{j,t}|$. I can test this prediction by

estimating a GLM very similar to the model in equation (4). In fact, I only need to substitute the first nonconstant regressor with the prediction error generated by the conformity mechanism, and estimate:

$$b^v(t) = \alpha + \beta_1^v I_{\mathrm{PD}}(t) \times |\, x_{j,t} - x_{i,t} \,| + \beta_2^v I_{\mathrm{PD}}(t) \times [(x_{j,t} - x_{j,t-1}) - (x_{i,t} - x_{i,t-1})] \\ + \beta_3^v I_{MKT}(t) \times (R_t\, x_{i,t}) + \beta_4^v controls + \varepsilon(t) \quad (5)$$

where the vector of controls is exactly the same as specified in equation (4) and described in section III.B. Contrary to the conformity hypothesis, I do not find any activity in the pre-specified vSt region that is significantly associated with the distance between peer investment allocations at the omnibus threshold of $p<0.05$ SVC. In summary, this null result casts some doubt on the ability of the conformity hypothesis to explain the observed peer effects.

IV. Discussion

In this paper I show how neural data can be used to test between competing mechanisms that generate peer effects. I find that neural activity in the vSt, an area known to encode changes in expected utility, can be decomposed into a social learning component and a relative wealth component. I then find that the cross-sectional variation in peer effects is explained by neural activity that is generated in response to seeing a peer's investment allocation. Together, these results provide novel evidence that peer effects are driven by a combination of a preference-based and a belief-based mechanism. The neural data are particularly useful in this setting for two reasons; first, while the social preference and social learning channels often make similar predictions about trading behavior, I show that they make distinct predictions at the neural level, allowing me to cleanly test for the contribution of each. Second, I use the neural data to construct a novel variable that can explain the cross-sectional variation in peer effects, and hence, in individual trading behavior.

It is important to highlight that the validity of many results here relies on the tight level of experimental control that the current laboratory setting affords. For example, one complication in testing peer effects mechanisms is due to selective communication where an investor tells others about his investment successes, but not about his failures (Kaustia and Knüpfer (2012); Han and Hirshleifer (2013)). This selection bias is shut down by design in the current experiment, as the investment decision of each subject in the *social treatment* is unconditionally sent to his peer. This allows me to show that, consistent with a relative wealth mechanism, an investor's utility is *decreasing* in his peer's investment return. Another valuable aspect of the laboratory setting is the random assignment combined with the ability to control for common shocks; as discussed above, this is sufficient to sidestep the well-known endogeneity problem in identifying peer effects (Manski (1993)).

While my results indicate that both a social learning mechanism and relative wealth mechanism contribute to the observed peer effects, one potential concern with this interpretation is that I assume each channel is acting in isolation. However, it is possible that there is a more complex relationship between the two channels that involves an interaction. In particular, I assume that the information contained in a peer's investment decision is valuable only for learning about the distribution of returns, but alternatively, this information can also be valuable for learning about a peer's investment skill (Boorman et al. (2013)). If this is the case, the subject is faced with a more complex learning problem where he must track both the asset's return distribution and his peer's level of skill. The current experimental design is not optimized to test this "dual-learning" problem because without several layers of additional assumptions, additional data on a subject's direct assessment of his peer's skill would be needed.

Another potential concern about the results shown here is the degree to which they are externally valid outside the lab. All data from this experiment are collected while subjects are trading inside an fMRI scanner, and like most lab experiments, the stakes are relatively low compared to those in the field. However, because the main objective of this study is to understand

the mechanisms that generate peer effects, it is not critical to match the quantitative effect sizes that are observed in the field. As long as substantial peer effects are observed in the lab data, just as they are in field data, this should provide a sufficient setting to probe the mechanisms using neural data. Indeed, this basic neurofinance methodology of replicating a robust trading behavior in the lab, and then using neural data to test existing theories can be applied to study other trading phenomena besides peer effects.

REFERENCES

Abel, Andrew B., 1990, Asset prices under habit formation and catching up with the joneses, *The American Economic Review* 38-42.

Ahern, Kenneth, Ran Duchin, and Tyler Shumway, 2014, Peer effects in risk aversion and trust, *Review of Financial Studies* Forthcoming.

Asch, Solomon E., 1951, Effects of group pressure upon the modification and distortion of judgments, *Groups, leadership, and men* 222-236.

Ashby, F. Gregory, 2011. *Statistical analysis of fmri data* (The MIT Press).

Beshears, John, James Choi, David Laibson, Brigitte Madrian, and Katherine Milkman, 2014, The effect of providing peer information on retirement savings decisions, *Journal of Finance* Forthcoming.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, 1992, A theory of fads, fashion, custom, and cultural change as informational cascades, *Journal of Political Economy* 992-1026.

Boorman, Erie D., John P. O'Doherty, Ralph Adolphs, and Antonio Rangel, 2013, The behavioral and neural mechanisms underlying the tracking of expertise, *Neuron* 80, 1558-1571.

Bossaerts, Peter, and Elise Payzan-Le Nestour, 2014, Learning about unstable, publicly unobservable payoffs, *Review of Financial Studies* Forthcoming.

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.

Brown, Jeffrey R., Zoran Ivković, Paul A. Smith, and Scott Weisbenner, 2008, Neighbors matter: Causal community effects and stock market participation, *Journal of Finance* 63, 1509-1531.

Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman, 2014, Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions, *Econometrica* 82, 1273-1301.

Caplin, Andrew, and Mark Dean, 2008, Dopamine, reward prediction error, and economics, *The Quarterly Journal of Economics* 123, 663-701.

Caplin, Andrew, Mark Dean, Paul W. Glimcher, and Robb B. Rutledge, 2010, Measuring beliefs and rewards: A neuroeconomic approach, *Quarterly Journal of Economics* 125, 923-960.

De Martino, Benedetto, John P. O'Doherty, Debajyoti Ray, Peter Bossaerts, and Colin Camerer, 2013, In the mind of the market: Theory of mind biases value computation during financial bubbles, *Neuron* 79, 1222-1231.

DeMarzo, Peter M., Ron Kaniel, and Ilan Kremer, 2008, Relative wealth concerns and financial bubbles, *Review of Financial Studies* 21, 19-50.

Duflo, Esther, and Emmanuel Saez, 2003, The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment, *Quarterly Journal of Economics* 815-842.

Eyster, Erik, and Matthew Rabin, 2010, Naive herding in rich-information settings, *American economic journal: microeconomics* 2, 221-243.

Fliessbach, Klaus, Bernd Weber, Peter Trautner, Thomas Dohmen, Uwe Sunde, Christian E. Elger, and Armin Falk, 2007, Social comparison affects reward-related brain activity in the human ventral striatum, *Science* 318, 1305-1308.

Frydman, Cary, Nicholas Barberis, Colin Camerer, Peter Bossaerts, and Antonio Rangel, 2014, Using neural data to test a theory of investor behavior: An application to realization utility, *Journal of Finance* 69, 907-946.

Frydman, Cary, and Colin Camerer, 2014, Neural measures of regret and reinforcement learning from investment experiences, *Working Paper*.

Gali, Jordi, 1994, Keeping up with the joneses: Consumption externalities, portfolio choice, and asset prices, *Journal of Money, Credit and Banking* 1-8.

Han, Bing, and David Hirshleifer, 2013, Self-enhancing transmission bias and active investing, *Working Paper*.

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.

Hirshleifer, David, and Siew Hong Teoh, 2009, Thought and behavior contagion in capital markets, *Handbook Of Financial Markets: Dynamics And Evolution, Handbooks in Finance* 1-46.

Hong, Harrison, Wenxi Jiang, Na Wang, and Bin Zhao, 2014, Trading for status, *Review of Financial Studies*.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, 2004, Social interaction and stock market participation, *Journal of Finance* 59, 137-163.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, 2005, Thy neighbor's portfolio: Word of mouth effects in the holdings and trades of money managers, *Journal of Finance* 60, 2801-2824.

Huettel, Scott, Allen Song, and Gregory McCarthy, 2004. *Functional magnetic resonance imaging* (Sinauer Associates).

Ivković, Zoran, and Scott Weisbenner, 2007, Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices, *Review of Financial Studies* 20, 1327-1357.

Kaustia, Markku, and Samuli Knüpfer, 2012, Peer performance and stock market entry, *Journal of Financial Economics* 104, 321-338.

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.

Kuhnen, Camelia, 2014, Asymmetric learning from financial information, *Journal of Finance* Forthcoming.

Kuhnen, Camelia M., and Brian Knutson, 2005, The neural basis of financial risk taking, *Neuron* 47, 763-770.

Lerner, Josh, and Ulrike Malmendier, 2013, With a little help from my (random) friends: Success and failure in post-business school entrepreneurship, *Review of Financial Studies* 2411-2452.

Lin, Alice, Ralph Adolphs, and Antonio Rangel, 2012, Social and monetary reward learning engage overlapping neural substrates, *Social Cognitive and Affective Neuroscience* 7, 274-281.

Lo, Andrew W., and Dmitry V. Repin, 2002, The psychophysiology of real-time financial risk processing, *Journal of Cognitive Neuroscience* 14, 323-339.

Lohrenz, Terry, Meghana Bhatt, Nathan Apple, and P. Read Montague, 2013, Keeping up with the joneses: Interpersonal prediction errors and the correlation of behavior in a tandem sequential choice task, *PLoS computational biology* 9.

Penny, William, Karl Friston, John Ashburner, Stefan Kiebel, and Thomas Nichols, 2006. *Statistical parametric mapping: The analysis of functional brain images* (Academic Press).

Pessiglione, Mathias, Ben Seymour, Guillaume Flandin, Raymond J. Dolan, and Chris D. Frith, 2006, Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans, *Nature* 442, 1042-1045.

Poldrack, Russell A., Jeanette Mumford, and Thomas Nichols, 2011. *Handbook of functional mri data analysis* (Cambridge University Press).

Pool, Veronika, Noah Stoffman, and Scott Yonker, 2014, The people in your neighborhood: Social interactions and mutual fund portfolios, *Journal of Finance* Forthcoming.

Preuschoff, Kerstin, Peter Bossaerts, and Steven R. Quartz, 2006, Neural differentiation of expected reward and risk in human subcortical structures, *Neuron* 51, 381-390.

Rutledge, Robb B., Mark Dean, Andrew Caplin, and Paul W. Glimcher, 2010, Testing the reward prediction error hypothesis with an axiomatic model, *The Journal of Neuroscience* 30, 13525-13536.

Schultz, Wolfram, Peter Dayan, and P. Read Montague, 1997, A neural substrate of prediction and reward, *Science* 275, 1593-1599.

Shue, Kelly, 2013, Executive networks and firm policies: Evidence from the random assignment of mba peers, *Review of Financial Studies* 26, 1401-1442.

Smith, Alec, Terry Lohrenz, Justin King, P. Read Montague, and Colin F. Camerer, 2014, Irrational exuberance and neural crash warning signals during endogenous experimental market bubbles, *Proceedings of the National Academy of Sciences* 111, 10503-10508.

Sokol-Hessner, Peter, Ming Hsu, Nina G. Curley, Mauricio R. Delgado, Colin F. Camerer, and Elizabeth A. Phelps, 2009, Thinking like a trader selectively reduces individuals' loss aversion, *Proceedings of the National Academy of Sciences* 106, 5035-5040.

Tversky, Amos, and Daniel Kahneman, 1974, Judgment under uncertainty: Heuristics and biases, *Science* 185, 1124-1131.

Table 1. Effects of peer investment and Sharpe ratio on risky asset allocation. Dependent variable is subject *i's* investment on trial *t*. *Peer Previous Investment* is subject *j's* investment on trial *t-1*. *Conditonal Sharpe* is the conditional Sharpe ratio using data from periods *{1,2,…,(t-1)}, Previous Investment* is subject *i's* investment on trial *t-1*. *Social* is a dummy that takes on the value 1 if the subject is in the *social treatment.* Standard errors are clustered by subject and *t*-statistics are in parentheses below estimated coefficients. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level respectively.

	(1) Social Treatment	(2) Control Condition	(3) Pooled
Peer Previous Investment	0.053 (2.50)**	0.015 (1.23)	0.144 (1.24)
Conditional Sharpe	0.003 (2.26)**	0.008 (3.80)***	0.008 (4.06)***
Previous Investment	0.597 (9.08)***	0.424 (9.72)***	0.508 (11.22)***
Social*Peer Previous Investment			0.063 (2.36)**
Social*Conditional Sharpe			-0.004 (-1.84)*
Constant	0.133 (5.85)***	0.194 (11.36)***	0.161 (9.85)***
Observations	4320	4320	8640
R^2	0.461	0.296	0.375
Subj. Fixed Effects	Y	Y	Y

Figure 1. Schematic of the three screens displayed in a typical trial in the fMRI experiment. In each of the two hundred trials in the experiment, subjects see a series of three screens, nearly identical to those shown below. In the first screen, subjects are asked to enter a fraction of their wealth to allocate to the risky asset. Subjects enter this fraction using a handheld device while inside the fMRI scanner, and are allowed to invest in increments of 0.1. On the second screen, subjects see the realized return of the risky asset, and their updated portfolio value as a function of the current period's risky asset allocation. Finally on the third screen, if the subject is in the *social treatment*, he sees his peer's risky asset allocation for that trial. If the subject is in the *control condition,* he sees a randomly generated number that is uniformly distributed over [0, 1].



Figure 2. BOLD measurements of neural activity. (A) Because fMRI measures the blood oxygen level dependent (BOLD) response, and not neural activity itself, a mapping from neural activity to the BOLD response is needed in order to make inferences about changes in neural activity. This mapping is known as the canonical hemodynamic response function, and is shown here as a function of one arbitrary unit of instantaneous neural activity at time 0. (B) This figure shows the BOLD response (the dashed line) that results from three sequential sources of neural activity (the solid line). The BOLD response combines linearly across multiple sources of neural activity.

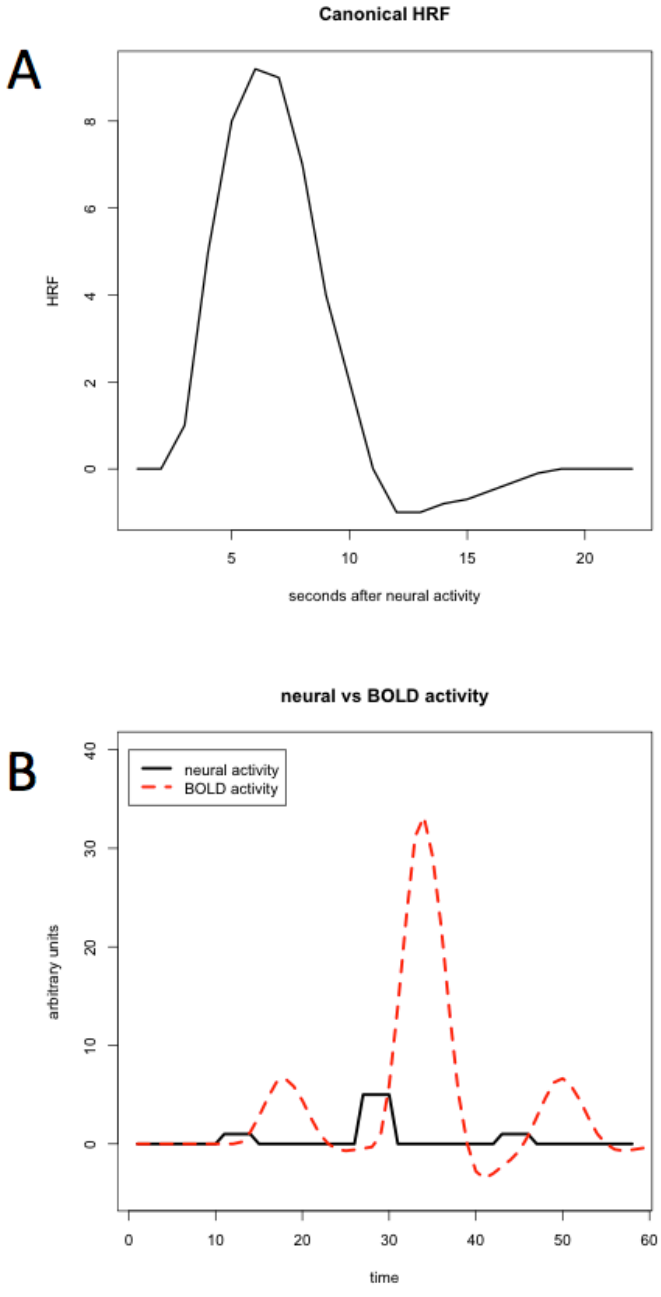


Figure 3. Subject-specific correlations between investment allocation and previous peer investment allocation. For each of the *i=1,…,24* subjects in the *social treatment*, two correlation measure are shown. The *within pair* correlation (blue) between subject *i's* investment allocation in period *t* and his peer's investment allocation in period *t-1* is computed as $WP(i) = corr(x_{i,t}, x_{j,t-1})$. The *across pair* correlation (red) represents the correlation between subject *i* in the *social treatment* and subject *k* in the *control condition.* It is computed as $AP(i) = \frac{1}{24}\sum_{k=1}^{24} corr(x_{i,t}, x_{k,t-1})$, for each of the *k=1,…,24* subjects in the *control condition*.

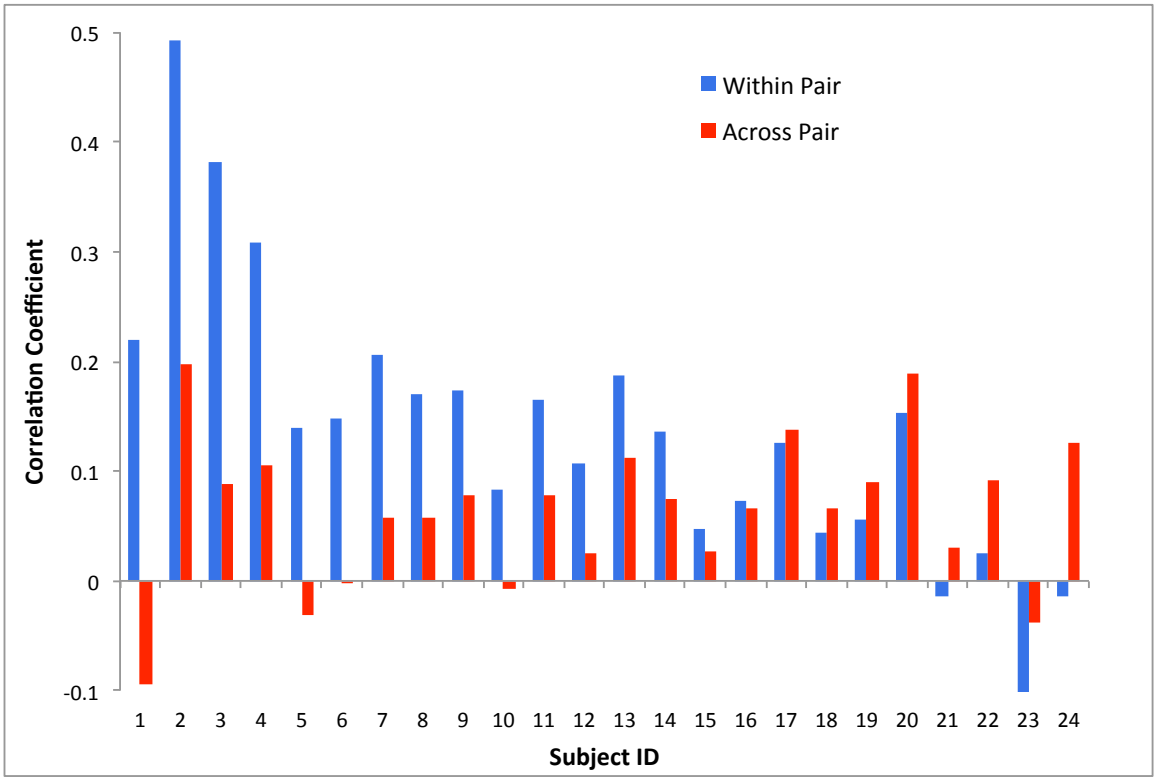


Figure 4. vSt reflects prediction errors generated from changes in wealth. The figure presents estimation results from equation (4):

$$b^v(t) = \alpha + \beta_1^v I_{PD}(t) \times R_t\left(x_{i,t} - x_{j,t}\right) + \beta_2^v I_{PD}(t) \times \left[\left(x_{j,t} - x_{j,t-1}\right) - \left(x_{i,t} - x_{i,t-1}\right)\right] +$$

$$\beta_3^v I_{MKT}(t) \times \left(R_t x_{i,t}\right) + \beta_4^v controls + \varepsilon(t)$$

Yellow and orange voxels are those that are in our pre-specified region of interest in the vSt. Red and orange voxels are those that exhibit activity at the time of a market screen onset that significantly correlate with the change in a subject's wealth that is generated from the realized market return. In other words, the voxels in red and orange are those for which we can reject the null hypothesis that $\beta_3 = 0$. All statistics are small volume corrected at p<0.05 using FWE. The $y = 16$ coordinate indicates the two-dimensional plane shown in the brain maps.

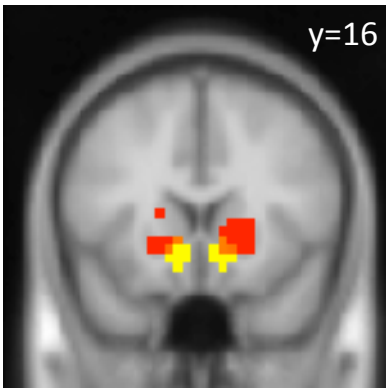


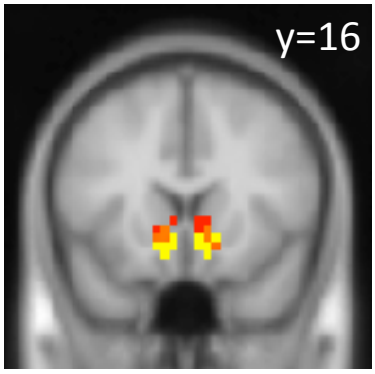
Figure 5. vSt reflects prediction errors generated under multiple peer effect mechanisms. The figure presents estimation results from equation (4):

$$b^v(t) = \alpha + \beta_1^v I_{PD}(t) \times R_t\left(x_{i,t} - x_{j,t}\right) + \beta_2^v I_{PD}(t) \times \left[\left(x_{j,t} - x_{j,t-1}\right) - \left(x_{i,t} - x_{i,t-1}\right)\right] +$$

$$\beta_3^v I_{MKT}(t) \times \left(R_t x_{i,t}\right) + \beta_4^v controls + \varepsilon(t)$$

Yellow and orange voxels are those that are in our pre-specified region of interest in the vSt. (A) Red and orange voxels are those that exhibit activity at the time of peer decision screen onset that significantly correlate with the social learning prediction error. These are voxels for which we can reject the null hypothesis that $\beta_2 = 0$. (B) Red and orange voxels are those that exhibit activity at the time of peer decision screen onset that significantly correlate with the relative wealth prediction error. These are voxels for which we can reject the null hypothesis that $\beta_1 = 0$. All statistics are small volume corrected at p<0.05 using FWE. The $y = 16$ coordinate indicates the two-dimensional plane shown in the brain maps.

A



B

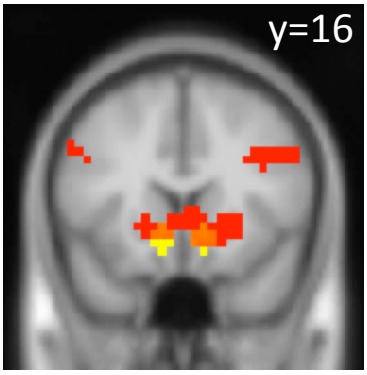


Figure 6. Correlation between brain activity and measures of peer effects.
Each of the twenty-four data points in the figure represents a single subject from the *social treament*. The degree to which vSt activity is correlated with the prediction errors generated by the social learning and relative wealth mechanisms, is itself correlated with the size of peer effects across subject.

