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

CARY FRYDMAN *

March 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 distinguish between competing preference-based explanations of peer effects. The observed neural activity is largely consistent with a preference for social status and relative wealth concerns. More generally, the neural data provide direct evidence that a change in relative wealth generates a utility shock that is distinct from a utility shock generated by a change in absolute wealth.

*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 or the outcome from a peer's decision may directly enter another agent's utility function (the preference channel). While there is typically only a single mechanism behind the information channel – belief updating via Bayes' rule – there are a variety of competing mechanisms within the class of preference-based explanations of peer effects. For example, an investor may be affected by his peer because of a concern for social status, an aversion to inequality, or because of a direct taste for conformity. Distinguishing between these mechanisms, particularly in a dynamic investment setting, can be difficult because direct measures of relative wealth concerns are difficult to obtain.

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 because it can

be used to directly measure a concern for relative wealth in a dynamic investment setting. Although survey data is also effective in measuring subjective well-being as a function of relative income (Luttmer (2005); Card et al. (2012)), neural data offers the advantage that it can be measured directly at the precise moment when relative wealth changes, without any potential biases driven by self-report. In addition to distinguishing between competing theories of peer effects, the neural data collected in this setting is also valuable as it allows for novel tests of the functional form of utility over relative wealth. This is important because the asset pricing predictions of models where an investor has a preference for status depend critically on the curvature of the investor's utility over relative wealth (Roussanov (2010)). The neural data allow me to run basic tests for concavity and convexity of this function by exploiting time series variation in relative wealth changes over the course of the fMRI experiment.

In addition to the advantages that the neural data offer in this setting, the trading data analyzed here are also useful in circumventing well-known obstacles in identifying causal peer effects. In particular, the trading data is collected in a highly controlled laboratory environment where subjects are randomly assigned into peer groups. This setup is attractive because it allows me to sidestep the 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)[1]. 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[2].

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 example, under the preference for status mechanism, I denote this quantity by the function *f(x),* and for the inequality aversion mechanism, I denote this quantity by the function *g(x).* The vector *x* contains the history of investment allocations and the current market return. The functions *f(x)* and *g(x)* therefore represent the prediction errors that are generated under the preference for status and inequality aversion mechanisms, 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 a subject has a preference for status, while a significant correlation between vSt activity and *g(x)* suggests that a subject has an aversion to inequality.

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

[1] See, for example, Schultz et al. (1997), Pessiglione et al. (2006), and Caplin et al. (2010).
[2] 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)).

separately identify whether a change in utility is driven by absolute changes in wealth or relative changes in wealth. Without the neural data, it is very difficult to distinguish between these two channels.

The three main results of the paper can be summarized as follows. First, by exploiting the random assignment of peers and the high level of experimental control in the lab environment, I find strong evidence of causal peer effects in individual investment decisions. Second, I use neural activity to test novel predictions of competing preference-based explanations of peer effects. The observed neural activity is largely consistent with a preference for social status, but is not consistent with a direct preference for conformity or an aversion to inequality. Testing between these competing theories without neural data is difficult because direct measures of relative wealth concerns are not readily available. Finally, I provide direct evidence that changes in relative wealth generate utility shocks that are distinct from those generated by changes in absolute wealth. Moreover, I show that in the current setting, the utility function over relative wealth is increasing, but concave. This suggests that the disutility from falling behind one's peer is larger than the utility gained from getting ahead of one's peer. On a methodological level, these results also provide an example of using a novel empirical framework 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 literatures on peer effects and relative wealth concerns. While recent empirical work has achieved growing success in cleanly identifying causal peer effects, there has been little work on investigating the mechanisms that drive these peer effects. One exception is a recent field experiment that uses a clever experimental design to separate an investor's revealed preference for a risky asset from the ownership of the risky asset (Bursztyn et al. (2014)). These authors can then infer whether investors are purchasing an asset due to social learning (information) or social utility (preferences). Bursztyn et al. (2014) find that

both mechanisms explain a portion of the observed peer effects, which is consistent with the neural results reported here[3].

My study builds on these results by distinguishing between competing theories of peer effects *within* the class of social utility mechanisms. While Bursztyn 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 direct evidence that when an asset delivers a negative return, a subject derives a negative utility shock if he invests 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 Bursztyn 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 peer effects in financial decision-making are strong enough to operate even in the absence of prior social interaction.

This paper also contributes to a vast literature on relative wealth concerns and preferences for status, starting with Duesenberry (1949). This literature is too large to survey here, but I highlight the works that are most relevant to my study. Within experimental economics, there is a large and rapidly growing literature showing that social preferences – such as a concern for status or an aversion to inequality – can explain many empirical regularities in experimental games (Falk et al. (2008); Heffetz and Frank (2010)). And while it is difficult to measure relative wealth concerns with field data, some empirical work has achieved success in demonstrating that survey measures of subjective well-being are negatively related to a peer's income (Luttmer (2005); Card et al. (2012)). Preferences for status, such as "Keeping up with the Jones'" or "Getting ahead of the Jones'" have also been used to explain empirical regularities in

[3] Goeree and Yariv (2015) use a lab experiment to disentangle social learning from conformity based explanations for herd behavior, but they do not test for a preference for status.

asset pricing and portfolio choice (Abel (1990); Campbell and Cochrane (1999); DeMarzo et al. (2008); Roussanov (2010)).

Methodologically, this paper 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)). There has also been a large amount of work in cognitive neuroscience that has investigated the neural mechanisms of social preferences (Fehr and Camerer (2007)). In perhaps the most closely related study to mine, Fliessbach et al. (2007) show that neural activity in the vSt is a function of social comparison, although their setting is one of effort provision and not of investment. In another related study, Tricomi et al. (2010) show that neural activity in the vSt is consistent with a direct concern for inequality aversion.

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 new methodology that can be used for testing competing mechanisms of economic behavior. I illustrate an example of this methodology by (i) using a theory from neuroscience to generate novel predictions of peer effects mechanisms, and then (ii) testing these novel predictions with neural data. 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 study uses vSt activity generated upon viewing stock returns to test whether regret theory can explain stock purchasing behavior (Frydman and Camerer (2015)).

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.

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 generates 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 between 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 or on the outcome of *j's* decision. This channel is referred to as the social preferences channel. Because all subjects are endowed with the same information at the beginning of the experiment, there is no reason that subject *i* should learn information about the asset from subject *j's* decision, and the social learning channel should not be active in this experiment. In what follows, I therefore focus on distinguishing between competing theories of peer effects within the class of preference-based explanations. In particular, I perform tests between a preference for status, inequality aversion, and a direct 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 ***A,*** I denote the change in expected discounted utility that is generated by news as the function *a(x);* this function represents the time series of prediction errors that is theoretically predicted by mechanism ***A.*** I can then test whether the time series of vSt activity significantly correlates with *a(x)*; if it does, this suggests that the mechanism used to generate the prediction errors – and summarized by the *a(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 that is generated by a preference for social status. Such preferences can arise due to relative wealth concerns such as "Keeping up with the Jones'" and "Getting Ahead of the Jones,'" or they may arise because subjects have a concern for status per se (Heffetz and Frank (2010)). Because the only channel through which status changes in the current experiment is through changes in relative wealth, I do not take a strong stance on the source of status preferences. However, it is important to note that subjects are not given any incentives that induce a preference over positional concerns or relative wealth: all subjects are paid exclusively as a function of their final portfolio value.

Under a preference for status, subjects derive utility from the difference between their peer's portfolio value and their own portfolio value[5]. 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 a preference for status 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 the social status mechanism is defined by, $f(\mathbf{x}) = R_t \times (x_{i,t} - x_{j,t})$.

Prediction 2 (Social status): If peer effects are driven by a preference for status, then at the time when a peer decision is revealed, vSt activity should positively correlate with the prediction error function defined by: $f(\mathbf{x}) = R_t \times (x_{i,t} - x_{j,t})$.

Next, I turn to Prediction 3, one that characterizes the neural activity that should be observed if peer effects are generated through an aversion to inequality (Rabin (1993); Fehr and Schmidt (1999)). I focus on a preference for outcome-based inequality aversion as opposed to intention-based inequality aversion, as intentions are likely to be less relevant in the current

[5] 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.

competitive market setting (Falk, Fehr and Fischbacher (2008)). The core feature of outcome-based inequality aversion models is that agents have an intrinsic preference against unequal outcomes, whether they are disadvantageous (an agent earns less than his peer) or advantageous (an agent earns more than his peer). In the current setting, these preferences imply that a subject's utility decreases in the distance between his payoff and his peer's payoff. Because a subject learns about his peer's payoff at the peer decision screen (after he has learned about his own payoff), a prediction error should be generated at the moment this information is revealed.

Specifically, a subject should derive a utility shock that is decreasing in the distance between his portfolio return and his partner's portfolio return. To be consistent with my tests for a preference for status, I use the difference in the percentage change in portfolio values instead of the difference in the dollar change in portfolio values. The prediction error generated under a preference for outcome-based inequality aversion should then be defined by, $g(\boldsymbol{x}) = |R_t \times (x_{i,t} - x_{j,t})|$. As this difference in payoffs grows, subject *i* should experience a larger utility decrease, and hence, the vSt should negatively correlate with this prediction error.

Prediction 3 (Inequality aversion): If peer effects are driven by the inequality aversion mechanism, then at the time when a peer decision is revealed, vSt activity should negatively correlate with the prediction error function defined by: $g(\boldsymbol{x}) = |R_t \times (x_{i,t} - x_{j,t})|$.

The final neural prediction specifies the vSt activity that is predicted to occur at the peer decision screen under an intrinsic taste for conformity (Asch (1951)). A large literature in social psychology indicates that people have a direct taste to follow others, and that they may derive disutility when their actions are different from those expressed in a peer group. In particular, a preference for conformity is likely to manifest itself in situations where there is no objectively correct action to take, thus causing an agent to use a peer's action as a social anchor on which to base his behavior.

In the current setting, if subject *i* is influenced by subject *j* because of an intrinsic taste to follow his peer, then on each trial subject *i* should derive a utility shock that is decreasing in the distance between his own risky asset allocation and that of subject *j's* risky asset allocation. This utility shock predicts that vSt activity of subject *i* should negatively correlate with $|x_{i,t} - x_{j,t}|$ at the moment when a peer decision is revealed. This leads to the final neural prediction:

Prediction 4 (Conformity): If peer effects are driven by the conformity mechanism, then at the moment when a peer decision is revealed, vSt activity should negatively correlate with the prediction error function defined by: $h(\mathbf{x}) = |x_{i,t} - x_{j,t}|$.

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 status or inequality aversion 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

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

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 $p<0.05$[6].

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 $k=1,\ldots,24$ subjects in the *control condition.* I then average these twenty-four 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[7]. Because the experimental design ensures that subjects in the *control condition* and *social treatment* observe

[6] 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 15-voxel extent threshold.
[7] 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.

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 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)[8]. Interestingly, when pooling data across control and treatment groups (column 3), the interaction between the

[8] 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.

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 anchoring drives peer effects, 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.

B. Neural Response to Market Returns

Before turning to the results on neural predictions 2-4, 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)

$$b^v(t) = \alpha + \beta_1^v I_{PD}(t)\times R_t\left(x_{i,t} - x_{j,t}\right) + \beta_2^v I_{PD}(t)\times\left|R_t\left(x_{i,t} - x_{j,t}\right)\right| + \beta_3^v I_{PD}(t)\times\left|x_{i,t} - x_{j,t}\right| + \beta_4^v I_{MKT}(t)\times\left(R_t x_{i,t}\right) + \beta_5^v controls + \varepsilon(t)$$

This is the same model that will be used to perform tests of predictions 2-4, 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 signed relative change in wealth that is revealed at the peer decision screen; the second nonconstant regressor is the absolute value of relative change in wealth. The third nonconstant regressor, $I_{PD}(t)\times\left|x_{i,t} - x_{j,t}\right|$, represents the conformity-based prediction error and the fourth 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) an indicator function denoting the onset of a peer decision screen

interacted with $(x_{i,t} - x_{j,t})$[9] and 7) 12 movement regressors that control for subject head movement inside the scanner.

Controls (1) - (6) are convolved with the HRF, whereas control (7) 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_4^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_4^v, averaged across subjects, is significantly positive ($p<0.05$ SVC). Because an increase in wealth should unambiguously increase expected discounted utility[10], this result acts to validate the neural methodology that I use to test the key neural predictions in the next section.

[9] This control is added to all GLMs because this quantity was found to significantly negatively correlate with vSt activity in the study by Lohrenz et al. (2013). Those authors describe this quantity as an "interpersonal prediction error," which is distinct from all three prediction errors I use to test the competing theories of peer effects.

[10] An alternative hypothesis is that the prediction error generated at the market screen is responding to the market return itself, independent of the news it carries about the change in wealth. However, as described above, the market return at the onset of the market screen is included in the vector of controls in this GLM, which is sufficient to rule out this alternative hypothesis.

C. Test of Neural Prediction 2

I now turn to the neural predictions that investigate the extent to which neural activity in the vSt reflects prediction errors that are generated under the three different preference-based 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 or survey data alone. I proceed in testing prediction 2 by performing a hypothesis test about the estimated coefficients from equation (4). Figure 5 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 8 voxels where β_1^v, averaged across subjects, is significantly positive ($p<0.05$ SVC)[11]. 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 preference for status mechanism; hence the observed vSt activity is consistent with prediction 2.

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

[11] This particular cluster is comprised of 65 voxels, but only 8 voxels lie within the vSt target region.

screen cannot be fully driven by changes in absolute wealth, because this information was revealed before the peer decision screen.

In addition to this logic of prediction error theory, it is useful to run additional empirical tests to rule this out alternative hypothesis. To do this, first note that under a preference for status, holding subject *i's* absolute wealth fixed, utility should increase when a peer's wealth decreases. This prediction can be tested using the neural data by estimating a new GLM that separately enters a peer's change in wealth as its own regressor:

(5)

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

where the controls vector is identical to the one specified in equation (4). The key test is whether β_2^v is significantly negative, which would indicate that, holding subject *i's* wealth fixed, subject *i's* utility is decreasing in subject *j's* wealth. To implement this test, for each subject I compute the average β_2^v across all 56 voxels in the vSt target region, and I denote this average by β_2^{vSt}. I then average this quantity across all twenty-four subjects in the *social treatment.* I find that the average $\beta_2^{vSt} = -0.1331$ *(p=0.06)*, and a similar computation reveals that $\beta_1^{vSt} = 0.1828$ *(p=0.02)*. β_2^{vSt} has the predicted negative sign, and is significantly different from zero at the 10% level. The fact that β_1^{vSt} is significantly positive suggests that a portion of the prediction error generated from the revelation of a peer's allocation is actually attributed to a change in absolute wealth.

This last result may seem to contradict the logic highlighted above that any utility shock generated by a change in absolute wealth should be reflected in neural activity measured at the "market" screen before the peer's allocation is revealed. However, if the relative wealth function

is nonlinear. then the prediction error generated at the peer decision screen may be a function of the change in absolute wealth. I investigate this potential nonlinearity in section III.E.

D. Test of Neural Predictions 3 and 4

In contrast to the result for prediction 2, I find no evidence consistent with prediction 3 or prediction 4. Specifically, using the estimation results from equation (4), even at a liberal threshold of $p<0.01$ uncorrected, I find no voxels in the vSt target region for which I can reject the null hypothesis that β_2^v is different from zero. Similarly, at the same liberal threshold , I find no voxels in the vSt target region for which I can reject the null hypothesis that β_3^v is different from zero. These null results indicate that at the time of the peer decision screen, vSt activity does not correlate with the theoretical prediction errors generated under the inequality aversion mechanism or the conformity mechanism.

The failure to reject the null hypotheses in this setting should be interpreted with some caution. In particular, the hypothesis tests I run assume that if any of the three preference mechanisms are present, then they act in isolation without any interaction between each other. It is possible that a subject has a concern for status but is also averse to inequality. When a subject experiences a negative relative wealth shock, both mechanisms predict a decrease in utility. However, when a subject experiences a positive relative wealth shock, the preference for status predicts an increase in utility while the aversion to inequality predicts a decrease in utility. This interaction will produce a utility over relative wealth that is nonlinear, which I investigate in the next section.

E. Functional Form of the Utility over Relative Wealth

The neural results from section III.C support the prediction that the utility function over relative wealth is increasing. However, it is possible that there may be nonlinearities in this function such that the utility loss from falling behind a peer may be larger than the utility gain

from moving ahead of a peer. This would imply that the relative wealth utility function exhibits a type of "social loss aversion." Alternatively, the function may be convex which occurs when the desire to "Get ahead of the Jones'" is stronger than the desire to "Keep up with the Jones'" (Roussanov (2010)). To distinguish between these to alternatives, I re-estimate a version of GLM (4) that partitions the change in relative wealth into its positive and negative components:

(6)

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

β_1^v represents the sensitivity of neural activity in voxel *v* to increases in relative wealth while β_2^v represents the sensitivity of neural activity in voxel *v* to decreases in relative wealth, at the peer decision screen. For each subject, I compute the average β_1^v and β_2^v across all 56 voxels in the vSt target region. I then average this quantity across all twenty-four subjects in the *social treatment.* I find that the average $\beta_1^{vSt} = 0.044$ *(p=0.30)* and the average $\beta_2^{vSt} = 0.089$ *(p=0.03)*. The sensitivity of the vSt to relative decreases in wealth is therefore twice as large as the sensitivity of the vSt to relative increases in wealth. This suggests that the relative wealth utility function is concave, and indicates that the disutility from falling behind one's peer is larger than the utility from getting ahead of one's peer.

The neural data can also provide information on whether specific types of relative wealth changes are more important than others. To see this, note that subject *i's* relative wealth increases in two circumstances: 1) when the market return is positive and subject *i* invested more than his peer and 2) when the market return is negative and subject *i* invested less than his peer. Conversely, subject *i's* relative wealth *decreases* in two circumstances: 1) when the market return is positive and subject *i* invested less than his peer and 2) when the market return is negative and

subject *i* invested more than his peer. To test whether each of these four circumstances contributes to a change in relative wealth in a similar manner, I first estimate the following GLM:

(7)

$$b^v(t) = \alpha + I_{PD}(t)\times[\beta_1^v I_{PosMktPosDiff}(t) + \beta_2^v I_{PosMktNegDiff}(t) + \beta_3^v I_{NegMktPosDiff}(t) + \beta_4^v I_{NegMktNegDiff}(t)] + \beta_5^v I_{MKT}(t) I_{PosMkt}(t) + \beta_6^v I_{MKT}(t) I_{NegMkt}(t) + \beta_7^v controls + \varepsilon(t)$$

where again, the controls vector is identical to the one used in equation (4) except I no longer include the market return at the time of the market revelation screen. The first four non-constant regressors are dummy variables that denote each of the four possible situations that can occur at the peer decision screen. For example, $I_{PosMktPosDiff}(t)$, takes on the value one when the there is a positive market return and a positive difference between the risky asset allocation of subject *i* and subject *j:* $(x_{i,t} - x_{j,t})>0$. In contrast, $I_{PosMktNegDiff}(t)$ takes on the value one when there is a positive market return and a negative difference between the risky asset allocation of subject *i* and subject *j,* $(x_{i,t} - x_{j,t})<0$. $I_{NegMktPosDiff}(t)$ and $I_{NegMktNegDiff}(t)$ are defined in an analogous manner. Finally, I also include dummy variables for positive and negative market returns at the time when market returns are revealed.

For each of the coefficients β_c^v, $c = 1,..,6$, I compute the average β_c^v across the 56 voxels in the vSt target region, yielding the estimate β_c^{vSt}. I then average β_c^{vSt} across all subjects in the *social treatment*, and display these results in Figure 6. The first two columns denote the average change in vSt activity at the time when the market return is revealed (β_5^{vSt} and β_6^{vSt}, respectively). The next four columns denote the average change in vSt activity at the time when the peer decision is revealed. For example, the fifth column denotes the average change in vSt activity at the moment when a subject learns his peer has invested more than him and there is a

positive market return. The observed significant decrease in vSt activity on these trials suggests there is a negative utility shock, consistent with the decrease in relative wealth. All six coefficients are significantly different from zero, except the coefficient that summarizes the change in vSt activity on trials where a subject invests less than his peer and the market delivers a negative return (the last column of Figure 6). In this circumstance, relative wealth increases, but the absence of a change in vSt activity does not reflect the associated positive utility shock. This asymmetry is consistent with the earlier result in this section that the utility function over relative wealth is concave.

IV. Discussion

In this paper I provide evidence of peer effects in individual trading behavior and then show how neural data can be used to test between competing mechanisms that generate these peer effects. I find that at the moment when a subject is informed about his peer's investment decision, an area of the brain known to encode changes in expected utility exhibits activity that is consistent with a preference for status. In particular, I find that even after information about a change in absolute wealth is revealed to a subject, the vSt responds strongly to information about a peer's decision insofar as it reflects changes in *relative* wealth. In contrast, the neural data do not support the two other competing preference-based explanations of a direct taste for conformity or an aversion to inequality.

This lack of evidence in favor of inequality aversion in the current setting does not suggest that inequality aversion is unimportant or inactive in other economic settings. Indeed, theoretical work has shown that inequality aversion preferences exhibit an important interaction with the economic environment such that these preferences are more likely to be expressed when agents have the ability to implement fair outcomes – which is not the case in the current experiment (Fehr and Schmidt (1999)). Other models of inequality aversion that are based on

reciprocity are also likely to be important in other settings where intentions can be more easily assessed (Rabin (1993)). With these caveats in mind, the data here suggest that peer effects in competitive financial market settings are more likely to be driven through a preference for status than through an aversion to inequality.

In addition to distinguishing between competing preference-based explanations of peer effects, the neural data used here is valuable in providing direct measures of relative wealth concerns. While there is a large body of research showing that survey measures are effective in measuring subjective-well being (Benjamin et al. (2014)) and that they can be used to infer relative wealth and income concerns (Luttmer (2005); Card et al. (2012)), I show that neural data can provide direct evidence that changes in relative wealth generate utility shocks that are distinct from those caused by changes in absolute wealth. Moreover, given the prevalence of relative wealth concern preferences in the finance literature (Abel 1990; Campbell and Cochrane (1999); Roussanov (2010)), my results can also be viewed as a validation of the assumption of relative wealth concerns in these models. Specifically, I show that utility is decreasing in a peer's wealth despite the fact that subjects are not endowed with an explicit incentive that induces this type of preference. This evidence is therefore most relevant to models that assume relative wealth concerns are an exogenous component of the utility function, compared to an endogenous one that may be generated by competition for scarce local resources (DeMarzo et al. (2008)).

It is important to highlight that many of my results rely 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 only 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. 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)).

Finally, one 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, just as they are in field, 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 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, *American Economic Review* 80, 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).

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Nichole Szembrot, 2014, Beyond happiness and satisfaction: Toward well-being indices based on stated preference, *American Economic Review* 104, 2698-2735.

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* 100, 992-1026.

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.

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205-251.

Caplin, Andrew, and Mark Dean, 2008, Dopamine, reward prediction error, and economics, *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.

Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez, 2012, Inequality at work: The effect of peer salaries on job satisfaction, *American Economic Review* 102, 2981-3003.

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.

Duesenberry, James S., 1949. *Income, saving, and the thoery of consumer behavior* (Harvard University Press, Cambridge, MA).

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* 118, 815-842.
Falk, Armin, Ernst Fehr, and Urs Fischbacher, 2008, Testing theories of fairness—intentions matter, *Games and Economic Behavior* 62, 287-303.
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 Klaus M. Schmidt, 1999, A theory of fairness, competition, and cooperation, *Quarterly Journal of Economics* 817-868.
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, 2015, Neural measures of regret and reinforcement learning from investment experiences, *Working Paper*.
Goeree, Jacob, and Leeat Yariv, 2015, Conformity in the lab, *Journal of the Economic Science Association,* forthcoming.
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.
Heffetz, Ori, and Robert Frank, 2010, Preferences for status: Evidence and economic implications, in Jess Benhabib, Alberto Bisin, and Matthew Jackson, eds.: (North Holland, Netherlands).
Hirshleifer, David, and Siew Hong Teoh, 2009, Thought and behavior contagion in capital markets, in Thorsten Hens, and Klaus Reiner Schenk-Hoppe, eds.: *Handbook of financial markets: Dynamics and evolution* (North Holland).
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.
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* 26, 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, e1003275.

Luttmer, Erzo F. P., 2005, Neighbors as negatives: Relative earnings and well-being, *Quarterly Journal of Economics* 963-1002.

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.

Rabin, Matthew, 1993, Incorporating fairness into game theory and economics, *American Economic Review* 1281-1302.

Roussanov, Nikolai, 2010, Diversification and its discontents: Idiosyncratic and entrepreneurial risk in the quest for social status, *Journal of Finance* 65, 1755-1788.

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.

Tricomi, Elizabeth, Antonio Rangel, Colin F. Camerer, and John P. O'Doherty, 2010, Neural evidence for inequality-averse social preferences, *Nature* 463, 1089-1091.

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*. *Conditional 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 10% increments. 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.

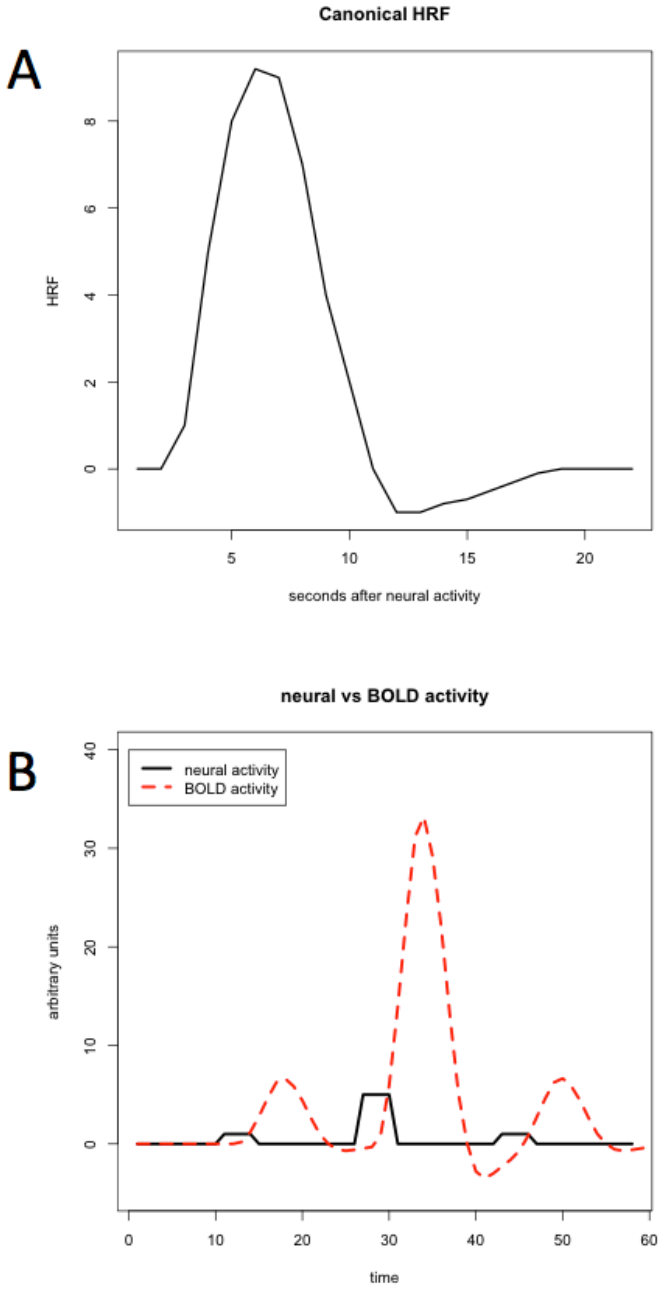


Figure 3. Subject-specific correlations between investment allocation and previous peer investment allocation. For each of the *i=1,…,24* subjects in the *social treatment*, two correlation measure are shown. The *within pair* correlation (blue) between subject *i's* investment allocation in period *t* 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*.

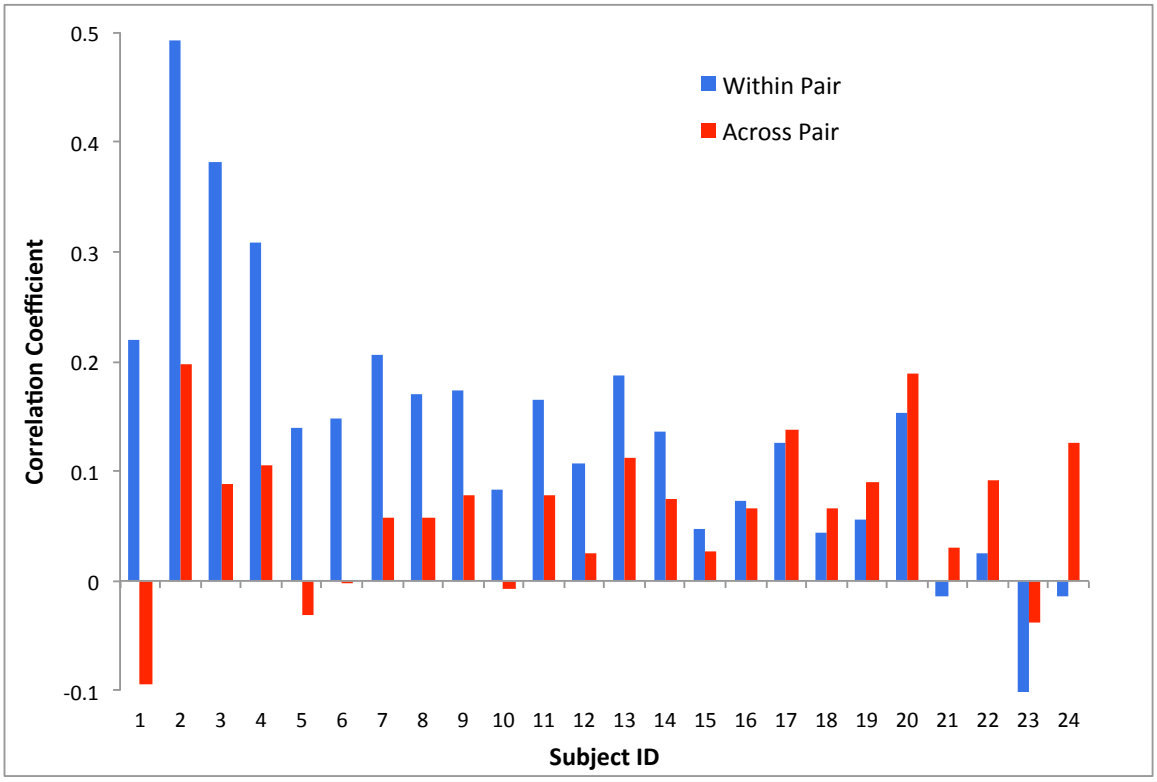


Figure 4. vSt reflects prediction errors generated from changes in absolute wealth. The figure presents estimation results from equation (4):

$$b^v(t) = \alpha + \beta_1^v I_{PD}(t) \times R_t\left(x_{i,t} - x_{j,t}\right) + \beta_2^v I_{PD}(t) \times \left|R_t\left(x_{i,t} - x_{j,t}\right)\right| + \beta_3^v I_{PD}(t) \times \left|x_{i,t} - x_{j,t}\right| +$$

$$\beta_4^v I_{MKT}(t) \times \left(R_t x_{i,t}\right) + \beta_5^v controls + \varepsilon(t)$$

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 $\beta_4 = 0$. 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.

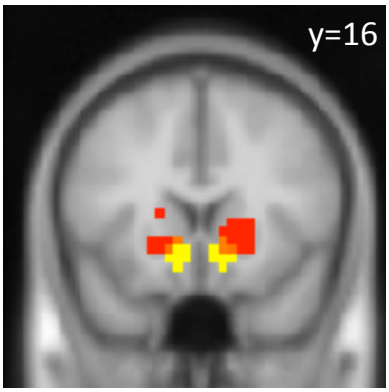


Figure 5. vSt reflects prediction errors generated under a preference for status. The figure presents estimation results from equation (4):

$$b^v(t) = \alpha + \beta_1^v I_{PD}(t) \times R_t\left(x_{i,t} - x_{j,t}\right) + \beta_2^v I_{PD}(t) \times \left|R_t\left(x_{i,t} - x_{j,t}\right)\right| + \beta_3^v I_{PD}(t) \times \left|x_{i,t} - x_{j,t}\right| +$$

$$\beta_4^v I_{MKT}(t) \times \left(R_t x_{i,t}\right) + \beta_5^v controls + \varepsilon(t)$$

Yellow and orange voxels are those that are in the pre-specified region of interest in the vSt. Red and orange voxels are those that exhibit activity at the time of peer decision screen onset that significantly correlate with the status preference prediction error. These are voxels for which I reject the null hypothesis that $\beta_2 = 0$, implying that these voxels exhibit activity that significantly positively correlates with a change in relative wealth at the moment when a peer decision is revealed. 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.

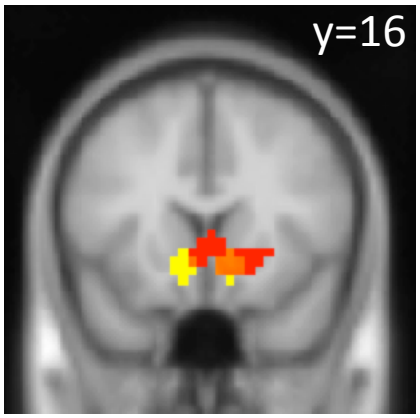


Figure 6. Direct measures of changes in utility as a function of changes in absolute and relative wealth. Each column displays the magnitude and standard error of the coefficient estimates β_1^{vSt} through β_6^{vSt} from equation (7):

$$b^v(t) = \alpha + I_{PD}(t)\times[\beta_1^v I_{PosMktPosDiff}(t) + \beta_2^v I_{PosMktNegDiff}(t) + \beta_3^v I_{NegMktPosDiff}(t) + \beta_4^v I_{NegMktNegDiff}(t)] + \beta_5^v I_{MKT}(t) I_{PosMkt}(t) + \beta_6^v I_{MKT}(t) I_{NegMkt}(t) + \beta_7^v controls + \varepsilon(t)$$

The first two columns represent the average change in vSt activity at the moment when a change in absolute wealth is revealed (at the market screen). The subsequent four columns represent the average change in vSt activity when a change in relative wealth is revealed (at the peer decision screen). *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level respectively.

