Herding and Contrarianism: A Matter of Preference?

Chad Kendall*

July 25, 2019

Abstract

Herding and contrarian strategies produce informational inefficiencies when investors ignore private information, instead following or bucking past trends. In a simple market model, I show theoretically that investors with prospect theory preferences generically follow herding or contrarian strategies, but do so because of future returns as opposed to past trends. I conduct a laboratory experiment to test the theory and to obtain an estimate of the distribution of preferences in the subject population. I find that approximately 70% of subjects have preferences that induce herding. Using the preference estimates, I quantify informational efficiencies and predict trade behavior in more general environments.

1 Introduction

Herding and momentum strategies, and their antithesis, contrarian strategies, interest financial economists because of their implications for the informational efficiency of market prices, which are in turn important due to potential effects on the real economy.¹,² What drives investors to use these strategies and to choose one over the other? The most basic explanation for herding/momentum (following the trend) is simply imitation, a form of ‘herd

---

¹Firm prices are important both for the allocation of capital and as signals to managers about their underlying investment decisions. See Bond, Edmans, and Goldstein (2012) for a recent review of the literature on the real effects of secondary financial markets.

²By a ‘momentum strategy’, I’m referring to the strategy of buying if prices have risen and selling if they’ve fallen, sometimes referred to as positive feedback trading (De Long (1990)), and not to be confused with the cross-sectional strategy of shorting losers and going long winners (which goes by the same name). Momentum and contrarian strategies are frequently discussed by practitioners and a small academic literature suggests different types of individuals or firms pursue each (see Grinblatt, Titman, and Wermers (1995), Grinblatt and Keloharju (2000), Baltzer, Jank, and Smajlbegovic (2015), and Grinblatt et al. (2016)).
mentality’ (Mackay (1841)), but behavioral explanations such as extrapolative expectations (De Long et al. (1990), Greenwood and Shleifer (2014), Barberis, Greenwood, Jin, and Shleifer (2015, 2016)) and rational explanations based on information externalities (Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992)) have also been theorized. Similarly, contrarian strategies (going against the trend) have rational explanations given a suitable informational environment (Avery and Zemsky (1998) and Park and Sabourian (2011)). In all of these explanations, the observation (or inference) of previous investor actions and/or prices is critical and, in fact, the strategies are often defined in terms of inference from past trends (Avery and Zemsky (1998) and Park and Sabourian (2011)). In this paper, I show instead that what appear to be strategies that depend upon past trends may in large part actually be independent of historical data. Rather, due to the nature of the relationship between past price trends and subsequent expected returns, these strategies may simply be driven by preferences over future returns.

I make my case for a preference-based explanation using both theory and a laboratory experiment. In the theoretical contribution, I revisit a standard model of trading with asymmetric information, that of Glosten and Milgrom (1985). In this model, herding and contrarianism are impossible with risk-neutral investors (Avery and Zemsky (1998)), but both behaviors are common in past experimental tests of the model (Cipriani and Guarino (2005, 2009) and Drehmann, Oechssler, and Roider (2005)). To resolve this apparent contradiction, I consider the behavior of an investor with cumulative prospect theory (CPT) preferences (Kahneman and Tversky (1992)). I show that such an investor generically (for almost all preference parameters) either buys or sells the tradable asset at extreme prices, independently of her private information. CPT preferences therefore provide a unifying explanation for herding and contrarianism. These behaviors are not, however, driven by a change in prices: investors’ decisions in the model are independent of history. Instead, decisions are driven by the fact that extreme prices imply highly skewed returns, and prospect theory gives preferences over skewness (Barberis and Huang (2008)).

My theoretical model predicts that CPT preferences can produce herding and contrarianism, but to what extent do they? Data from past experiments (Cipriani and Guarino (2005, 2009) and Drehmann, Oechssler, and Roider (2005)) is difficult to interpret because preferences, belief-based errors, as well as beliefs about others’ preferences and errors all contribute to behavior. I therefore conduct a new laboratory experiment in which I isolate the role of preferences. I find that herding and contrarianism are pervasive, suggesting that

---

3Momentum and herding produce similar trade patterns, but herding is often defined in the context of asymmetric information environments. See Devenow and Welch (1996) and Hirshleifer and Teoh (2003) for surveys of herding in financial markets.
CPT preferences may be a significant contributor to what appear to be trend-dependent strategies.

In the theoretical model, investors arrive sequentially to a market, trading a single, binary-valued asset with a market maker who posts separate bid and ask prices. Each investor, after receiving a private, binary signal about the asset’s fundamental value, may buy or sell a single unit of the asset (or abstain from trading). The standard result with risk-neutral, expected utility investors (Avery and Zemsky (1998)) is that each investor trades according to her private information - buying with a favorable signal and selling otherwise. With CPT preferences, I show that an investor ignores her private information at extreme prices, instead trading in a single direction. The direction of trade depends upon the relative strengths of utility curvature and probability weighting. In fact, with power utility and a single parameter probability weighting function, the direction of trade depends simply on the difference of the two parameters that govern these functions.

Probability weighting and utility curvature together imply preferences over skewness which in turn drive trading behavior. At a high price, the expected return from buying the asset is negatively skewed: it is very likely to return a small positive amount (when the asset is valuable), but occasionally results in a large loss (when it is not). Selling the asset instead results in a positively skewed return. As has been recognized (Barberis and Huang (2008)), the overweighting of small probabilities in CPT generates a preference for positive skewness. Thus, when probability weighting dominates utility curvature, we observe contrarian behavior: selling at high prices. On the other hand, when utility curvature dominates, an investor exhibits a preference for negative skewness, leading to herding: buying at high prices. The opposing effects of probability weighting and utility curvature are relatively well understood within the literature that studies prospect theory preferences (e.g. Glimcher and Fehr (2013)), but have received less attention in the applied literature (with the exception of Barberis (2012)). Here, their opposing effects are clearly illustrated through the resulting simple analytic expressions that govern behavior.

Guided by the theoretical predictions, I conduct a laboratory experiment to assess the extent to which we should expect trend-dependent strategies to result from preferences, something which would be very difficult to determine in naturally-occurring settings. Critically, I shut down all social causes of behavior by having subjects make decisions in an individual decision-making environment. However, I maintain the framing of a market: subjects observe a past price trend and make buy or sell decisions. Doing so allows me to observe decisions when choices are framed in a natural way and to test for any direct effect of trends on decision-making.

In the first (MAIN) treatment, I control for Bayesian errors by providing subjects with the
correct Bayesian posteriors corresponding to their information sets (including both public and private information), allowing me to isolate the effect of preferences. By having subjects make a large number of trading decisions (60), I am able to estimate the distribution of preferences in the population and use it to assess (via simulation) the informational efficiency of markets composed of this population. In the second (NO BELIEFS) treatment, I determine whether belief formation errors exacerbate herding and contrarian behavior by keeping everything identical to the MAIN treatment except for the provision of correct Bayesian posteriors.

I find strong evidence of herding and contrarianism in both treatments, establishing a significant role for preferences in generating these behaviors. Moreover, comparing the two treatments, I find that Bayesian errors actually reduce herding leaving the level of contrarian behavior unchanged. In the MAIN treatment, where Bayesian errors are absent, approximately three-quarters of subjects exhibit either herding or contrarian behavior and, among these, 90% follow herding strategies. I find only weak evidence that the nature of the trend itself affects decisions.

Estimating preference parameters using data from the MAIN treatment, I find that 75% of subjects are better described by CPT preferences than expected utility (CRRA). Importantly, this result does not depend on CPT having more degrees of freedom (a general critique of prospect theory), given that I took an ex ante stance on the reference point (the status quo) and establish theoretically that (under the assumed functional forms) only the difference between the degree of utility curvature and probability weighting determines behavior. In fact, CPT better describes behavior than expected utility even when CPT is restricted to a single degree of freedom. Of the minority of subjects that are not best described by CPT, most are risk-neutral expected utility maximizers (which is technically consistent with either theory), but a few abstain from trading at extreme prices. This latter behavior cannot be reconciled with CPT preferences (but is consistent with a high level of risk-aversion), illustrating that, while the theory captures most subjects’ behavior, it is falsifiable.

I perform two exercises with the estimated distribution of preferences. First, I simulate markets composed of a sequence of investors (as in the theory) to compare informational efficiency relative to a benchmark market comprised of only risk-neutral investors. The simulations demonstrate not only that herding and contrarianism produce significant informational inefficiencies relative to the benchmark, but also that the relative composition of these behaviors is important. Both reduce informational efficiency by masking private information, but contrarian trades produce a higher variance in the inefficiencies observed. Second, assuming that the distribution of preferences is the same across treatments, I evaluate the types of Bayesian errors subjects make, on average. At intermediate priors, they underweight private information, while at extreme priors they underweight the information contained in
Lastly, I take the estimated preferences of the modal subject in the experiment (who exhibits herding behavior) and ask how an investor with these preferences would behave when facing actual market returns. Because market returns become more (less) skewed as prices rise (fall), a fact first documented by Chen, Hong and Stein (2011), if preferences over skewness drive behavior as in the theory and experiment, we should expect this investor to be willing to purchase an asset after rising prices even with negative private information, and to sell with positive information after falling prices. The simulation confirms this conjecture, demonstrating that preferences can drive herding in environments more general than the simple, binary asset value setup of the theory and experiment.

The seminal paper on herding in financial markets is that of Avery and Zemsky (1998) who show that herding and contrarianism are impossible with risk-neutral investors (unless additional sources of uncertainty are added to the model). More recent theoretical papers have studied whether or not non-expected utility preferences can generate these behaviors. Qin (2015) shows that regret aversion generates herding, but not contrarian, behavior. Boortz (2016) builds on Ford (2013), to show that ambiguity can generate both behaviors, but only with preferences that vary with the state. In contrast to these models, I show that static CPT preferences generate both types of behavior. Other papers have considered the possibility that hedging motives (Decamps and Lovo, 2008) or private values (Cipriani and Guarino, 2008) can induce herd behavior. The model here, instead, explains how herding and contrarianism can arise under common values and without hedging motives - in an environment identical to that of previous experiments in which these behaviors have been observed.

The paper also contributes to the experimental literature on herding in financial markets (see Cipriani and Guarino (2005), Drehmann, Oechssler, and Roider (2005), Cipriani and Guarino (2008), Cipriani and Guarino (2009), Park and Sgroi (2012), and Bisiere, Decamps, and Lovo (2015)). Bisiere, Decamps, and Lovo (2015) introduced the key methodological innovation of converting the market into an individual decision problem in order to study how preferences affect the revelation of private information. However, their study differs in that they focus on an environment in which hedging motives can cause standard risk preferences to generate herding and contrarianism. I find these behaviors even in an environment without hedging motives and put forward prospect theory as an explanation, a theory they do not consider.4

4In two robustness treatments, Bisiere, Decamps, and Lovo (2015)) eliminate the hedging motive. They continue to observe herding and contrarianism, inconsistent with risk-averse expected utility maximization but consistent with the model here.
Lastly, this paper contributes to a growing literature that applies CPT preferences to understanding investor behavior in financial markets. Several papers study the disposition effect, the tendency to sell recent winners but hang onto recent losers (Barberis and Xiong (2009), Barberis and Xiong (2012), Ingersoll and Jin (2013), Li and Yang (2012), Meng and Weng (2016)). Barberis and Huang (2008) study the pricing of securities when investors have CPT preferences, Barberis, Huang, and Thaler (2006) use loss aversion to explain stock market non-participation, and Levy, De Giorgi, and Hens (2012) and Ingersoll (2016) study CAPM under CPT preferences. Although not directly related to financial markets, Barberis (2012) is a closely related paper that shows how CPT preferences can explain the popularity of casino gambling.

2 Theory

2.1 Model

The model is a sequential trading model based on that of Glosten and Milgrom (1985). In each period \( t = 1, 2, \ldots, T \), a single new investor arrives to the market to trade an asset of unknown value, \( V \in \{0, 1\} \). I denote the initial prior that the asset is worth 1 by \( p_1 \in (0, 1) \). Upon arrival, an investor may either buy or sell short a single unit, or not trade, \( a_t \in \{\text{buy, sell, NT}\} \). After making her decision, the investor leaves the market. All trades are with a risk-neutral market maker who is assumed to face perfect competition, earning zero profit in expectation.\(^5\) The market maker incorporates the equilibrium informational content of the current order in setting prices. Specifically, he posts an ask price, \( A_t \), at which he is willing to sell a unit of stock and a bid price, \( B_t \), at which he is willing to buy a unit. When the asset value is realized at \( T \), investors who purchased the asset at time \( t \) receive a payoff of \( V - A_t \) and those who sold receive a payoff of \( B_t - V \) (no discounting). All market participants observe the complete history of trades and prices, denoted \( H_t = \{a_1, a_2, \ldots, a_{t-1}\} \cup \{A_1, A_2, \ldots, A_{t-1}\} \cup \{B_1, B_2, \ldots, B_{t-1}\} \).

Investors are one of three types: risk-neutral, prospect theory, or uninformed investors. Uninformed investors, who arrive with probability \( 1 - \mu, \mu \in (0, 1) \), trade for exogenous reasons and are equally likely to buy or sell (such traders are sometimes referred to as ‘noise’ traders).\(^6\) Risk-neutral investors have standard risk-neutral expected utility preferences and arrive with probability, \( \mu \gamma, \gamma \in (0, 1) \). Finally, prospect theory investors have the CPT preferences of Kahneman and Tversky (1992) (described in detail in Section 2.4) and arrive

\(^5\)This assumption is standard in the literature and follows Glosten and Milgrom (1985).

\(^6\)Assuming an equal likelihood of buying and selling is inconsequential and is easily relaxed.
with the remaining probability, \( \mu(1 - \gamma) \). The risk-neutral and CPT investors receive private information upon arrival to the market: a binary signal, \( s_t \in \{0, 1\} \), which has the correct realization with probability \( q = Pr(s_t = 1|V = 1) = Pr(s_t = 0|V = 0) \in (\frac{1}{2}, 1) \). All signals are independent conditional on \( V \). I refer to \( s_t = 1 \) as a favorable signal, and \( s_t = 0 \) as unfavorable.

I include two types of informed investors (risk-neutral and CPT) in order to contrast their behaviors, and because we would expect heterogeneous preferences in any population.\(^7\) Limiting to a single CPT preference type is for ease of exposition only and when I simulate markets based on the estimated distribution of preferences from the experiment (Section 4), I allow for many CPT types. Risk-neutral investors are not necessary for the main conclusions of the model and I discuss the model’s implications when they are not present.

### 2.2 Solution Concept

Being a game of asymmetric information, the solution concept is Perfect Bayesian Equilibrium. An equilibrium consists of a specification of the strategies of the risk-neutral and CPT investors, along with the bid and ask prices of the market maker. These prices depend upon the market maker’s knowledge of the distribution of preference types as well as his beliefs about their strategies. As usual, these beliefs, which are pinned down at every history due to the presence of the uniformed investors, must be correct in equilibrium. Strategies are functions of the complete history of prices and trades, as well as one’s private signal, to an action: buy, sell, or not trade. As these details are standard, I omit formal definitions.

### 2.3 Risk-Neutral and Uninformed Investors

The roles of the risk-neutral and uninformed investors, as well as the market maker, are standard. I describe them first before discussing the more novel behavior of the CPT investors.

In each period, the market maker posts separate bid and ask prices given by \( B_t = Pr(V = 1|H_t, a_t = sell) \) and \( A_t = Pr(V = 1|H_t, a_t = buy) \), respectively, a consequence of the assumption of perfect competition. Intuitively, the ask price exceeds the public belief, denoted \( p_t = Pr(V = 1|H_t) \), because a buy decision reflects favorable private information, \( s_t = 1 \), in equilibrium. Similarly, the public belief exceeds the bid price, resulting in the standard bid-ask spread, \( A_t - B_t > 0 \). Importantly, uninformed investors allow the adverse selection problem between informed investors and the uninformed market maker to be overcome.

\(^7\)Bruhin, Fehr-Duda, and Epper (2010), for example, argue that both expected utility and prospect theory preferences should be included in applied theoretical work because they find, as I do, a mixture of these preferences among their experimental subjects.
Due to their presence, the bid and ask prices do not fully reflect the private information of informed investors, who are then able to make profitable trades. The market maker loses money to informed investors, but recoups it from uninformed investors. This intuition is formalized in Lemma 1, which characterizes the behavior of the risk-neutral investors, showing that the standard result of Glosten and Milgrom (1985) continues to hold even in the presence of CPT investors. All proofs are provided in Appendix A.

**Lemma 1:** In any equilibrium, for all \( p_t \in (0, 1) \), risk-neutral investors always trade: those with favorable signals \( (s_t = 1) \) buy and those with unfavorable signals \( (s_t = 0) \) sell.

An immediate consequence of Lemma 1 is that, provided risk-neutral investors arrive with positive probability, some information is partially revealed in every period: an information cascade in which prices stagnate and subsequent trades reveal no new information never occurs. Thus, by the law of large numbers, public beliefs and bid and ask prices converge to the true asset value in the limit as \( T \to \infty \), achieving full informational efficiency. In Section 2.6, I contrast this result with the case in which only CPT investors exist.

### 2.4 Prospect Theory Investors

CPT differs from expected utility in that investors evaluate gains and losses relative to a reference point. Perhaps the simplest possible reference point is status quo wealth, which is the reference point I adopt.\(^8\) The behavioral asset pricing literature has tended to instead use the expected wealth from investing in a risk-free asset (see Barberis and Huang (2008), Barberis and Xiong (2009), and Li and Yang (2013)). In the absence of a risk-free asset, as is the case here, these different specifications are equivalent in the sense that the reference point is the amount an investor can attain without risk. Expectations-based reference points, such as those in Koszegi and Rabin (2006,2007) are another popular alternative. However, I show in Appendix D that they are generally inconsistent with the experimental evidence that follows.\(^9,10\)

---

\(^8\)Baillon, Bleichrodt, and Spinu (2019) show experimentally that the status quo and ‘maxmin’ reference points are the most common among many possible alternatives, including expectations-based reference points. They define the maxmin reference point as the maximum of the minimum outcomes across gambles in a choice set. In the environment here, it corresponds to the status quo.

\(^9\)I’m implicitly assuming investors evaluate their gains or losses when the asset value is realized, either by closing their position so that the gains or losses are realized (corresponding to the realization utility of Shefrin and Statman (1985)) or by evaluating their gains or losses on paper. In the experiment, this assumption is satisfied. Barberis and Xiong (2009) discuss the difference between paper gains and losses and realization utility, showing that the distinction can be important in a model in which investors make multiple trading decisions.

\(^10\)The issue of narrow or broad framing (Barberis, Huang, and Thaler (2006)) does not play a role in the model because only a single asset is available. With multiple assets or other sources of background
CPT specifies value functions, \( v^+() \) and \( v^-() \), and decision weight functions, \( w^+() \) and \( w^-() \), over gains and losses, respectively. The decision weight functions apply to capacities, a generalization of probabilities, but for binary outcomes result in simple non-linear transformations of the objective probabilities. The utility a CPT investor derives from a binary lottery, \( \mathcal{L} \), which returns a gain of \( x \) with probability \( r \) and a loss of \( y \) with probability \( 1-r \) is given by \( U(\mathcal{L}) = w^+(r)v^+(x) + w^-(1-r)v^-(y) \).\(^{11}\)

Given this utility function, a CPT investor with a private belief, \( b_t = Pr(V = 1|H_t, s_t) \), prefers buying to not trading if

\[
w^+(b_t)v^+(1-A_t) + w^-(1-b_t)v^-(A_t) \geq 0
\]

where the utility of not trading results in no gain or loss and is normalized to zero. Similarly, she prefers selling to not trading if

\[
w^+(1-b_t)v^+(B_t) + w^-(b_t)v^-(B_t - 1) \geq 0
\]

If neither equation (1) nor equation (2) is satisfied, then a CPT investor abstains from trading.

The forms of equations (1) and (2) are sufficiently general that little can be said about the behavior of the investor without imposing additional structure. I proceed using the functional forms for the value and decision weighting functions provided in the original work of Kahneman and Tversky (1992), because they are tractable, parsimonious, and appear to fit decisions over binary gambles reasonably well.\(^{12}\) Specifically, I assume

\[
\begin{align*}
v^+(x) &= x^\alpha \\
v^-(y) &= -\lambda(-y)^\alpha
\end{align*}
\]

and

\[
w^+(r) = w^-(r) = \frac{r^\delta}{(r^\delta + (1-r)^\delta)^{\frac{1}{\delta}}}
\]

\(^{11}\)See Kahneman and Tversky (1992) for the more general formulation for any number of outcomes, as well as an axiomatic foundation for the preferences.

\(^{12}\)Other functional forms, especially for the decision weighting function, have appeared in the literature. See Bruhin, Fehr-Duda, and Epper (2010) and the references therein.
Figure 1: Examples of the Value and Probability Weighting Functions

Note: Value function (left graph) and probability weighting function (right graph) for the case of $\alpha = 0.88$, $\lambda = 2.25$, and $\delta = 0.65$ (taken from the median estimates in Kahneman and Tversky (1992) and averaging the probability weighting parameters they separately estimate for gains and losses).

With $\alpha \in (0, 1]$, $\lambda \geq 1$, and $\delta \in (0, 1]$, $\alpha \in (0, 1)$ reflects the common experimental finding of risk-aversion over gains and risk-seeking over losses (an “S-shaped” value function). $\lambda \geq 1$ reflects loss-aversion: losses are weighted more heavily than gains. Finally, $\delta \in (0, 1]$ implies that low-probability events are overweighted. Figure 1 illustrates examples of each function.

Substituting the functional forms into equations (1) and (2) results in the following optimal decisions for a CPT investor:

\[
\begin{align*}
\text{buy if } & \quad \left( \frac{b_t}{1-b_t} \right)^\delta \geq \lambda \left( \frac{A_t}{1-A_t} \right)^\alpha \\
\text{sell if } & \quad \left( \frac{b_t}{1-b_t} \right)^\delta \leq \frac{1}{\lambda} \left( \frac{B_t}{1-B_t} \right)^\alpha
\end{align*}
\]

(3)

Risk-neutral investors are a special case of CPT investors with $\alpha = \delta = \lambda = 1$. Under this parameterization, equations (3) state that an investor buys when her belief exceeds the bid price and sells when her belief is below the ask price as in Lemma 1. More generally, we must explicitly evaluate the beliefs and prices. An investor with a favorable signal, $s_t = 1$, has a private belief conditional on the history and her private signal (denoted $b^1_t$) given by Bayes’ rule:

\[
b^1_t = \frac{p_t q}{p_t q + (1 - p_t)(1 - q)}
\]

\textit{Kahneman and Tversky assume a slightly more general form allowing }$w^+(r)$\textit{ and }$w^-(r)$\textit{ to have different parameters, but their experimental estimates for the two parameters are quantitatively similar. I assume a common parameter, which results in a significant increase in tractability.}
Similarly, an investor with an unfavorable signal, \( s_t = 0 \), has private belief (denoted \( b_t^0 \)):

\[
b_t^0 = \frac{p_t(1-q)}{p_t(1-q) + (1-p_t)q}
\]

The bid and ask prices can also be written as functions of the public belief. Because of perfect competition, Bertrand competition ensures the ask price is the minimum price at which the market maker earns zero expected profit. Similarly, the bid price is the maximum such price.

\[
A_t = \min \left\{ A_t \in [0, 1] : A_t = \frac{p_t Pr(a_t=buy|V=1)}{p_t Pr(a_t=buy|V=1)+(1-p_t)Pr(a_t=buy|V=0)} \right\}
\]

\[
B_t = \max \left\{ B_t \in [0, 1] : B_t = \frac{p_t Pr(a_t=sell|V=1)}{p_t Pr(a_t=sell|V=1)+(1-p_t)Pr(a_t=sell|V=0)} \right\}
\]

where the conditional probabilities of observing a purchase or a sale depend upon the proportion of risk-neutral and CPT investors, as well as their equilibrium strategies.

After substituting the expressions for private beliefs and bid and ask prices, the optimal decision of a CPT investor with a favorable signal becomes

\[
\begin{align*}
\text{buy if } & \left( \frac{p_t}{1-p_t} \right)^{\delta-\alpha} \geq \lambda \left( \frac{1-q}{q} \right)^{\delta} \left( \frac{Pr(a_t=buy|V=1)}{Pr(a_t=buy|V=0)} \right)^{\alpha} \\
\text{sell if } & \left( \frac{p_t}{1-p_t} \right)^{\delta-\alpha} \leq \frac{1}{\lambda} \left( \frac{1-q}{q} \right)^{\delta} \left( \frac{Pr(a_t=sell|V=1)}{Pr(a_t=sell|V=0)} \right)^{\alpha}
\end{align*}
\]

The corresponding equations for an investor with an unfavorable signal are identical except that the ratio of \( 1-q \) to \( q \) on the right-hand side is inverted in each.

Although the opposing effects of \( \alpha \) and \( \delta \) have received relatively little attention in applications of prospect theory (with the exception of Barberis’ (2012) model of casino gambling), it is immediately clear from (5) that their difference, \( \delta - \alpha \), plays a critical role. To understand the intuition, consider a simplified example. Remove all private information so that the bid and ask prices collapse to the public belief, \( p_t \). In this case, risk-neutral investors have no incentive to trade because their private beliefs correspond to that of the public belief (equal to price): the gambles corresponding to a purchase or a sale have zero expected value. With this simplification, equations (3) for a CPT investor become

\[
\begin{align*}
\text{buy if } & \left( \frac{p_t}{1-p_t} \right)^{\delta-\alpha} \geq \lambda \\
\text{sell if } & \left( \frac{p_t}{1-p_t} \right)^{\delta-\alpha} \leq \frac{1}{\lambda}
\end{align*}
\]

We see that, for public beliefs sufficiently different from one-half, either buying or selling is strictly preferred to not trading.

Consider a public belief, \( p_t > \frac{1}{2} \). As the decision weights become more distorted from
linearity ($\delta$ decreases from one), the propensity to buy decreases and the propensity to sell increases. Intuitively, a decrease in $\delta$ increases the weight assigned to the small probability, $1 - p_t$, of a loss and reduces the weight assigned to the larger probability, $p_t$, of a gain, thereby making buying less attractive. Conversely, it increases the utility from selling because the small probability is instead associated with a gain. Therefore, for $p_t$ sufficiently large and $\delta < \alpha$, the investor strictly prefers to sell the stock: she exhibits a preference for positive skewness, the consequence of prospect theory studied extensively in Barberis and Huang (2008).

Conversely, consider an increase in the curvature of the value function (decrease in $\alpha$ from one). Mathematically, we see that we get exactly the opposite effect from that due to an increase in the distortion of probabilities. Intuitively, as the curvature increases, the small gain $(1 - p_t)$ that occurs with probability $p_t$ if one buys is preferred to the large gain $(p_t)$ that occurs with probability $1 - p_t$ if one sells, a simple consequence of risk-aversion (an investor with risk-neutral preferences would be indifferent). At the same time, the small probability of a large loss if one buys is preferred to the large probability of a small loss if one sells, due to risk-seeking. Both effects make buying preferable to selling so that for $p_t$ sufficiently large and $\delta > \alpha$, the investor always strictly prefers to buy: she exhibits a preference for negative skewness.$^{14}$

Finally, consider the role of loss aversion. An increase in $\lambda$ reduces the range of public beliefs at which an investor is willing to trade, because it simultaneously makes each inequality in (6) more difficult to satisfy. The intuition here is simple: an increase in loss-aversion increases the disutility of losses, which makes one more likely to abstain from taking on a position in the asset. Perhaps surprisingly, however, loss aversion prevents trading only at intermediate public beliefs. Although the potential losses are larger at extreme public beliefs, one can take the side of the trade that either minimizes the probability ($\delta < \alpha$) or the size ($\delta > \alpha$) of a loss. At intermediate beliefs, on the other hand, the chance of a loss of medium size and probability can only be avoided by abstaining from trade.

This simple example captures the countervailing forces of distortions due to decision weights and utility curvature. The intuition carries over to the full equilibrium characterization I pursue in the following section.

$^{14}$Beyond the intuition, we can see that the preference to trade in a particular direction at an extreme price must be driven by the skewness of returns, and not by their mean or variance. For example, with a preference for negative skewness, a CPT investor switches from always selling to always buying as the price increases from 0 to 1. Over this range, the mean return is always positive if she trades with her private signal, so that, as in the case of a risk-neutral investor, an investor that cares only about the mean always trades with her private signal. The variance of the return increases from zero to its maximum value at $p = \frac{1}{2}$, and then decreases to zero again, so that if the variance in returns were driving behavior, we’d expect behavior to be the same at both price extremes.
2.5 Equilibrium

The equilibrium strategies of investors are functions of the public belief and their private signals. They are given by Lemma 1 for risk-neutral investors, and follow from equations (5) for CPT investors with a favorable signal and the corresponding equations for those with unfavorable signals. Note first the symmetry of the environment: an investor with a favorable signal at a public belief, \( p_t \), is in a symmetric situation to an investor with an unfavorable signal at a public belief, \( 1 - p_t \). Therefore, equilibrium behavior is symmetric around \( p_t = \frac{1}{2} \), simplifying its description.

As discussed in the previous section, absent private information, skewness preferences cause CPT investors to strictly prefer to trade in a particular direction. As prices become extreme, because the skewness in returns is unbounded, eventually skewness preferences dominate private information, causing such an investor to trade in the same direction for both realizations of her private signal. Thus, when the public belief is sufficiently large and \( \delta > \alpha \), CPT investors buy regardless of their private signal, and, for sufficiently small public beliefs, they always sell. Conversely, when \( \delta < \alpha \), CPT investors sell regardless of their private signal when the public belief is sufficiently large, and buy when the public belief is sufficiently small.

For less extreme public beliefs, CPT investors either trade according to their private information or abstain from trading. The public beliefs at which behavior transitions depend upon an investor’s private signal, so that an equilibrium is characterized by four transitions as illustrated in Figure 2. For the case of \( \delta > \alpha \), as the public belief increases, the CPT investor with a favorable signal is the first to transition from selling to abstaining. At this price, which I denote \( p^0 \), the equilibrium abruptly changes from a pooling equilibrium in which no information is revealed by a sale, to a separating equilibrium in which the CPT investor with a favorable signal abstains with probability one.\(^{15}\) As the public belief increases further, the CPT investor with a favorable signal then gradually transitions from abstaining to buying over the price range, denoted \( p^1 \equiv (\underline{p}^1, \overline{p}^1) \). This transition is gradual because the more the investor buys, the more information is revealed, which in turn raises the ask price. In equilibrium, the CPT with a favorable signal mixes. For the case of \( \delta < \alpha \), \( p^0 \) and \( p^1 \) are defined similarly, but the roles of investors with favorable and unfavorable signals are reversed (see Figure 2).

\(^{15}\)The zero profit condition holds at two bid prices for prices just below \( p^0 \). In the equilibrium case of a higher bid price, the CPT investor with a favorable signal sells with probability one. In the other case, she abstains with probability one, lowering the bid price. But, in this latter case, there exists a higher bid price at which the market maker can make strictly positive profits, selling to the CPT investor with the unfavorable signal. The assumption of perfect competition among market makers rules out this possibility.
Figure 2: Prospect Theory Investor Behavior in Equilibrium

Note: Behavior of prospect theory investors in equilibrium. The upper two plots correspond to $\delta > \alpha$, and the bottom two plots to $\delta < \alpha$. The left two plots illustrate a parameterization for which investors do not trade with either signal over some intermediate range of public beliefs. The right two plots illustrate a second parameterization in which investors instead trade according to private information over this range. Within a transition region, the distance between No Trade and Buy or Sell reflects the mixing probability in the mixed strategy equilibrium (a higher probability is associated with the closer action).

In Figure 2, the upper two plots correspond to $\delta > \alpha$ and the lower two to $\delta < \alpha$. Within each of these two cases, I illustrate the two possible relationships between the locations of the transitions. For the plots on the left of Figure 2, the parameters are such that the two transitions lie on opposite sides of $p = \frac{1}{2}$. In this case, neither type of investor trades over some range of intermediate beliefs. The plots on the right of Figure 2 illustrate a second parameterization in which the transitions lie on the same side of $p = \frac{1}{2}$. In this case, for intermediate public beliefs, a separating equilibrium exists in which CPT investors’ trades reveal their private information. Theorem 1 is the main theorem of the paper formalizing the illustration of Figure 2. For all parameterizations, the equilibrium is essentially unique (up to indifference at transition prices).
Theorem 1: In equilibrium:

1. The market maker posts bid and ask prices (given by (4)) where the conditional buy and sell probabilities are determined by the equilibrium strategies of informed investors that follow.
2. For all \( p_t \in (0, 1) \), risk-neutral investors buy with favorable signals and sell with unfavorable signals.
3. CPT investors’ strategies are as follows:
   (a) If \( \delta = \alpha \), there exist two cutoff values of loss aversion, \( \lambda > \Lambda > 1 \) such that, at all \( p_t \in (0, 1) \), if \( \lambda \leq \lambda \), CPT investors buy with favorable signals and sell with unfavorable signals, and, if \( \lambda \geq \lambda \), they do not trade with either signal. If \( \lambda \in (\Lambda, \Lambda) \), CPT investors mix between buying and not trading with favorable signals and between selling and not trading with unfavorable signals.
   (b) If \( \delta \neq \alpha \), for all \( \alpha \in (0, 1] \) and \( \delta \in (0, 1] \), there exist four transitions, \( p^0 \), \( p^1 \equiv (p^1_1, p^1_1) \), and their symmetric counterparts, that characterize strategies:
      i. If \( \delta > \alpha \), CPT investors with favorable signals sell for \( p_t \leq p^0 \), don’t trade for \( p_t \in (p^0, p^1) \), and buy for \( p_t \geq p^1 \). CPT investors with unfavorable signals sell for \( p_t \leq 1 - p^1 \), don’t trade for \( p_t \in (1 - p^1, 1 - p^0) \), and buy for \( p_t \geq 1 - p^0 \).
      ii. If \( \delta < \alpha \), CPT investors with unfavorable signals buy for \( p_t \leq p^0 \), don’t trade for \( p_t \in (p^0, p^1) \), and sell for \( p_t \geq p^1 \). CPT investors with favorable signals buy for \( p_t \leq 1 - p^1 \), don’t trade for \( p_t \in (1 - p^1, 1 - p^0) \), and sell for \( p_t \geq 1 - p^0 \).
      iii. Within each of the transitions, \( p^1 \) and \( 1 - p^1 \), CPT investors mix between abstaining and trading.
      iv. Transitions do not overlap: \( p^0 < p^1 \) and \( p^1 < 1 - p^0 \), implying \( p^0 < \frac{1}{2} \).

2.6 Equilibrium Properties

In the model, CPT investors buy or sell independently of their private signals as prices become extreme. By the original definition (Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992)), buying independently of one’s private signal qualifies as herding. However, by more recent definitions of herding and contrarianism in financial markets (Avery and Zemsky (1998) and Park and Sabourian (2011)), CPT investor behavior in the model would not necessarily qualify as herding or contrarianism. In these definitions, behaviors depend on the history of past actions. For example, an investor is said to ‘herd’ if she would sell at
some initial price (based upon her private information), but instead buys after others’ trades increase prices. A CPT investor that faces an extreme price instead exhibits behavior that is independent of the past history. If she buys at a particular price regardless of her signal, she does so whether the price rose (after a series of purchases) or fell from a more extreme price (after a series of sales). Of course, extreme prices are more likely to occur after a sequence of purchases which is why her behavior will tend to look like it depends upon past trends.

Given multiple definitions in the literature, depending upon one’s taste, CPT investor behavior could be defined as herding and contrarian or not. I have chosen to do the former in order to avoid further confusion through the introduction of new terms, providing precise definitions in Definition 1 (where I also define the term ‘unresponsive’ to encompass both).

Definition 1:

1. An informed investor exhibits herding behavior at the public belief, $p_t$, if independently of her signal, (i) $p_t > \frac{1}{2}$ and she buys, or (ii) $p_t < \frac{1}{2}$ and she sells.

2. An informed investor exhibits contrarian behavior at the public belief, $p_t$, if independently of her signal, (i) $p_t > \frac{1}{2}$ and she sells, or (ii) $p_t < \frac{1}{2}$ and she buys.

3. An informed investor exhibits unresponsive behavior at the public belief, $p_t$, if her behavior is either herding or contrarian.

Figure 2 and Theorem 1 show that CPT investors exhibit herding ($\delta > \alpha$) or contrarian ($\delta < \alpha$) behavior generically (unless $\delta = \alpha$) at sufficiently extreme public beliefs. In addition, the theory predicts ‘partial’ versions of these behaviors, defined as follows:

Definition 2:

1. An informed investor exhibits partial herding behavior at the public belief, $p_t$, if (i) $p_t > \frac{1}{2}$ and she buys with a favorable signal but abstains with an unfavorable signal, or (ii) $p_t < \frac{1}{2}$ and she sells with an unfavorable signal but abstains with a favorable signal.

2. An informed investor exhibits partial contrarian behavior at the public belief, $p_t$, if (i) $p_t > \frac{1}{2}$ and she sells with an unfavorable signal but abstains with a favorable signal, or (ii) $p_t < \frac{1}{2}$ and she buys with a favorable signal but abstains with an unfavorable signal.

As discussed in Section 2.3, the presence of risk-neutral investors in the model ensures prices achieve full information efficiency as $T \to \infty$. Conversely, in their absence, information cascades in which no information is ever revealed to the market can develop in one of two ways. First, if the prior is intermediate ($p_1 = \frac{1}{2}$) and the CPT investors are loss averse, they do not trade - a form of non-participation. Second, at extreme prices CPT investors
trade unresponsively so that no information is revealed. In both cases, the CPT investor’s decision leaves the price unchanged so that all subsequent investors face the same decision. Their decisions then also reveal no information, and so on. With both risk-neutral and CPT investors these two effects slow, rather than halt, information revelation, an effect which I quantify in Section 4.

3 Experiment

3.1 Design

The theoretical model predicts herding and contrarian behaviors even in the absence of social factors, something which would be difficult to ascertain in naturally-occurring settings. Previous controlled laboratory studies have found evidence of both. Table 2 of Cipriani and Guarino (2009), for example, reports that, as prices become extreme, approximately 20% of trades are herding trades, and 30-40% are contrarian. Even data from these past experiments is difficult to interpret, however. Subjects’ preferences, their abilities to form Bayesian posteriors, and, perhaps most importantly, beliefs about the strategies of other subjects, may all influence their decisions. Is the high frequency of contrarian behavior driven by a preference for positive skewness? Or, is it a best response to the belief that other subjects are herding due to their preferences for negative skewness? Although it would be possible to disentangle these explanations using a structural model, doing so would require strong assumptions about the beliefs subjects have about other subjects’ strategies. Fortunately, Bisiere, Decamps, and Lovo (2015) develop a more direct way to assess the role of preferences in these environments, using an individual decision-making task, an approach which I adopt.

In the individual decision-making task of my experiment, subjects choose between a safe option (not trading) and two risky options (buying or selling). In designing the task, I faced a design choice about how to frame these choices, which may be important. In the design, I depart from Bisiere, Decamps, and Lovo (2015) in two important ways. First, in

---

16Cipriani and Guarino (2005) and Drehmann, Oechssler, and Roider (2005) report lower levels of herding, but have low power to detect such behavior. Herding can only be detected in their experiments when a subject’s private signal contradicts the price trend, which occurs relatively infrequently. Cipriani and Guarino (2009) use the strategy method (as do I) to avoid this censoring issue.

17In a metastudy of herding in non-market settings, Weizsäcker (2010) shows that subjects do not have rational expectations about the behavior of others, so rational expectations, although a natural assumption, may not be a valid one.

18Kahneman and Tversky (1981) provide some of the first evidence of framing effects, but there are now many such examples in the literature. In the context of financial decisions, see, for example, Apesteguia, Oechssler, and Weidenholzer (2018).
line with real markets and financial market herding experiments with sequential trading, I provide subjects with a historical price path, whereas Bisiere, Decamps, and Lovo (2015) do not. In doing so, I hope to make my results more comparable to those of previous market experiments and to real-life investing decisions. In addition, by not eliminating the price paths altogether, I can explicitly test for effects of price paths not predicted by the theory. Second, I keep the framing across treatments identical - the only difference between my two treatments is whether or not subjects are given correct Bayesian beliefs conditional on their information sets. In their pair of main treatments, Bisiere, Decamps, and Lovo (2015) also compare providing correct Bayesian beliefs to not doing so, but they change the framing from a market to a lottery when providing correct beliefs.\textsuperscript{19}

More specifically, in the experiment I represent asset values as urns with 7 balls of one color and 3 of another ($q = 0.7$) as in Figure 3. Starting from a uniform prior, a subject observes a history of prices determined by a sequence of public signal draws (draws are with replacement). Subjects are explicitly told that the price represents the expected value of the asset (what it is worth on average) conditional on this public information. They participate in 30 such ‘markets’ consisting of 15 price paths and their symmetric counterparts. The price paths were chosen such that subjects make multiple decisions at the same price but with different histories, allowing explicit tests for history dependence (Appendix E provides the complete list of price paths). After observing each price path, a subject decides whether to buy, sell, or abstain for each possible realization of her private signal so that unresponsive (herding or contrarian) behavior can be directly observed (as in Cipriani and Guarino (2009)). After a subject makes her decision, her private signal is drawn, her corresponding trade is executed, and she receives feedback about both her realized signal, the asset value, and the resulting payoff. Subjects are paid for the decisions that correspond to their realized private signals in each market.

In the ‘MAIN’ treatment, subjects are given the Bayesian posteriors for each possible realization of their private signal, thus precluding Bayesian errors from driving behavior and allowing me to cleanly estimate preferences from their decisions. To determine the extent to which Bayesian errors contribute to herding and contrarianism, in the ‘NO BELIEFS’ treatment, I don’t provide Bayesian posteriors.

In both treatments, I follow the previous experimental literature (Cipriani and Guarino (2005) and Drehmann, Oechssler, and Roider (2005)) by having the market maker in the experiment (the experimentalist) post only a single price equal to the expected value of the

\textsuperscript{19}Bisiere, Decamps, and Lovo (2015) conduct a robustness treatment (SME) where they keep the framing of a market when providing correct beliefs. However, when analyzing Bayesian errors, they use their original treatments which potentially confound the effect of correct beliefs with a framing effect.
asset, $p_t$, rather than separate bid and ask prices. This procedure both simplifies the problem for subjects, and, more importantly, strengthens incentives to trade according to private information by making the difference between private beliefs and prices larger. Importantly, the predictions of the model are unchanged when trades occur at a single price, except that all transitions in behavior occur at degenerate threshold prices, guaranteeing unique predictions (see Appendix B for details).

I recruited undergraduate subjects at the University of California, Santa Barbara over the month of August, 2016. They participated in a computerized implementation of the experiment implemented in oTree (Chen, Schonger, and Wickens (2016). I conducted three sessions of each treatment for a total of 46 subjects in each. Each subject took part in 30 consecutive markets of a single treatment (MAIN or NO BELIEFS), making a total of 60 trading decisions (one for each possible realization of their private signal in each market). Subjects were paid for their decisions in each market with average earnings of $17.13. The experiment typically finished in just over an hour. The full set of instructions is provided in Appendix F.

### 3.2 Experimental Results

In Section 3.2.1, I present the raw experimental data to assess the extent to which preferences and Bayesian errors are responsible for herding and contrarianism. I also show that particular
### Table 1: Behavior by Treatment

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Abstain</th>
<th>Revealing</th>
<th>Herding</th>
<th>Contrarian</th>
<th>Partial Herding</th>
<th>Partial Contrarian</th>
<th>Irrational</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAIN</td>
<td>5.5</td>
<td>27.5</td>
<td>34.8</td>
<td>10.6</td>
<td>10.8</td>
<td>3.8</td>
<td>7.0</td>
</tr>
<tr>
<td>NO BELIEFS</td>
<td>5.2</td>
<td>37.4</td>
<td>13.9</td>
<td>12.8</td>
<td>9.6</td>
<td>16.1</td>
<td>5.2</td>
</tr>
</tbody>
</table>

Note: Percentages of each type of behavior in the MAIN treatment (Bayesian posteriors provided to subjects) and the NO BELIEFS treatment.

patterns predicted by the theory are present in the data. In Section 3.2.2, I show that CPT preferences provide a better fit to the data than expected utility, and present estimates of the distribution of preference parameters in the experimental population. Lastly, Section 3.2.3 assesses the Bayesian errors subjects make in the NO BELIEFS treatment.

### 3.2.1 The Role of Preferences

In the data analysis, I decided ex ante to drop the first 3 markets during which subjects are becoming familiar with the interface and environment.\(^{20}\) I begin by categorizing trading decisions in Table 1. In addition to the definitions of Definition 1 and Definition 2, I define behavior as ‘Revealing’ if a subject buys with a favorable and sells with an unfavorable signal, thus fully revealing her signal. I define ‘Abstain’ as not trading with either signal. Lastly, I define all other trading combinations as ‘Irrational’ given that they can’t be reconciled with the theory (or any other theory I’m aware of).\(^{21}\)

The results of Table 1 clearly demonstrate that social factors are not necessary for unresponsive behavior: in each treatment, more than 25% of behavior is herding or contrarian and less than 40% is consistent with the predictions of risk-neutral subjects (revealing trades). Furthermore, providing subjects with the correct Bayesian posteriors does not reduce unresponsive behavior as we’d expect if Bayesian errors drive this behavior. In fact, the opposite occurs - with correct Bayesian beliefs in the MAIN treatment, we observe significantly more herding (t-test, \(p = 0.000\)), with corresponding reductions in revealing (t-test, \(p = 0.000\)) and partial contrarian (t-test, \(p = 0.000\)) behaviors.\(^{22,23}\) Thus, belief errors appear to work

---

\(^{20}\)Including this data does not affect any of the qualitative results, nor does restricting the analysis to only the second half of the data. Learning seems to play a very limited role: the results of classifying individual subjects in Section 3.2.2 are remarkably similar when using only the first or second halves of the data.

\(^{21}\)Irrational behavior consists of trading against both signals, always buying or always selling at \(p_t = 0.5\), buying with an unfavorable signal but abstaining with a favorable signal, or selling with a favorable signal but abstaining with an unfavorable signal.

\(^{22}\)Bisiere, Decamps, and Lovo (2015) find a similar result when they compare their ME treatment (which requires Bayesian updating) to either of their LE or SME treatments (which do not).

\(^{23}\)In a particular form of erroneous Bayesian reasoning, extrapolative expectations (De Long et al. (1990), Greenwood and Shleifer (2014), Barberis, Greenwood, Jin, and Shleifer (2015,2016)), people extrapolate
against a preference-based tendency to herd, something which I explore further in Section 3.2.2.

**Result 1:** *Unresponsive behavior is common when social factors are absent. Bayesian errors are not solely responsible for this behavior: providing the correct Bayesian posteriors leads to an increase in herding.*

The results suggest Bayesian errors are not the sole cause of unresponsive behavior, provided subjects actually form beliefs and use them to make trading decisions. If they instead use simple heuristics, we might expect the precise nature of the historical price path to influence behavior. To test for this possibility, I ask whether or not the frequencies of herding or contrarianism differ when the price path is (i) monotonic (no contradictory public signals), or (ii) either monotonic or ends in a ‘streak’ (two or more public signals in the same direction). To do so, I define a ‘normalized’ price, a measure of the extremeness of the price which treats rising and falling prices symmetrically.\(^{24}\) I then run logit regressions of the behavior of interest (herding or contrarian) on a dummy variable indicating either monotonicity or a streak and normalized price fixed effects, restricting to observations at which unresponsive behavior is possible (normalized prices other than one-half). In each of the four regressions (\{herding,contrarian\} × \{monotonic,streak\}), I cluster standard errors at the subject level.\(^{25}\)

For herding, I find no statistical evidence of path-dependence in either treatment for either monotonic price paths or streaks (\(p > 0.59\) across the four specifications). Allowing herding to include partial herding doesn’t change this result (\(p > 0.167\) across the four specifications). On the other hand, I find some evidence that contrarianism is driven by the nature of the price paths. Monotonic price paths in the MAIN treatment, and streaks in both treatments, significantly reduce the frequency of contrarianism (\(p < 0.032\)), suggesting that subjects find it less desirable to go against the trend in these cases.\(^{26}\) The sizes of these future market returns from past returns. Under extrapolative expectations, we might expect subjects’ beliefs to be too extreme when not provided with the correct Bayesian posteriors, leading to more herding in the NO beliefs treatment, rather than less, as observed. Thus, the data here seems inconsistent with extrapolative expectations. However, note that the fact that the data-generating process is known here may limit the possibility of forming extrapolative expectations.

\(^{24}\)Formally, the normalized price is defined as \(p_t\) if \(p_t \geq \frac{1}{2}\) and \(1 - p_t\) if \(p_t < \frac{1}{2}\). Subjects do not always treat rising and falling prices symmetrically (see Section 3.2.2). Therefore, the results should be interpreted as average effects over rising and falling prices.

\(^{25}\)The normalized price affects unresponsive behavior theoretically and must be included because of mechanical correlations between price paths that are monotonic or have streaks and more extreme prices. Similarly, without restricting the regression to prices other than one-half, unresponsive behavior is mechanically related to both monotonicity and a streak: at a price of one-half, unresponsive behavior cannot occur (by definition) and no monotonic price paths or those ending in a streak result in this price in the data.

\(^{26}\)The sign is consistent with monotonic price paths reducing contrarianism in the NO BELIEFS treatment
Table 2: Behavior by Price

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Normalized Price</th>
<th>Abstain</th>
<th>Revealing</th>
<th>Herding</th>
<th>Contrarian Partial Herding</th>
<th>Partial Contrarian</th>
<th>Irrational</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAIN</td>
<td>0.50</td>
<td>2.2</td>
<td>61.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>36.3</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>2.2</td>
<td>39.7</td>
<td>18.5</td>
<td>10.3</td>
<td>20.9</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>0.84</td>
<td>5.7</td>
<td>19.6</td>
<td>44.9</td>
<td>12.5</td>
<td>9.8</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>0.93</td>
<td>8.0</td>
<td>10.1</td>
<td>55.4</td>
<td>12.7</td>
<td>6.2</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>0.97</td>
<td>15.2</td>
<td>12.0</td>
<td>50.0</td>
<td>14.1</td>
<td>4.4</td>
<td>2.2</td>
</tr>
<tr>
<td>NO BELIEFS</td>
<td>0.50</td>
<td>12.3</td>
<td>63.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>24.6</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>4.9</td>
<td>43.8</td>
<td>10.0</td>
<td>13.0</td>
<td>12.8</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>0.84</td>
<td>3.8</td>
<td>32.3</td>
<td>13.9</td>
<td>14.4</td>
<td>13.6</td>
<td>19.3</td>
</tr>
<tr>
<td></td>
<td>0.93</td>
<td>4.0</td>
<td>27.2</td>
<td>23.1</td>
<td>16.3</td>
<td>6.9</td>
<td>20.3</td>
</tr>
<tr>
<td></td>
<td>0.97</td>
<td>4.4</td>
<td>23.9</td>
<td>21.7</td>
<td>14.1</td>
<td>3.3</td>
<td>29.4</td>
</tr>
</tbody>
</table>

Note: Percentages of each type of behavior at a given normalized price. The normalized price pools high and low prices together (e.g. 0.7 refers to an actual price of 0.7 or 0.3).

effects are moderate, however, ranging from a 3.5 to 5% decrease in the probability of a contrarian trade. The evidence is consistent, therefore, with preferences being the primary determinant of behavior.

**Result 2:** Herding is not more frequent when price paths are monotonic or end with two or more public signals in the same direction. Contrarian behavior is less likely under both conditions, but only moderately so.

The theory predicts not only unresponsive behavior itself, but also particular patterns of this behavior with respect to prices (Figure 2). Specifically, revealing trades and abstention should be more common at intermediate prices and herding and contrarianism more common at extreme prices. To test these predictions, Table 2 breaks down the frequency of each type of behavior at each normalized price.

In both treatments, the frequency of revealing behavior almost perfectly monotonically decreases, while the frequency of herding increases, as predicted. I confirm these results via logit regressions of a dummy variable indicating the type of behavior on the normalized price, clustering standard errors at the subject level. For the herding regressions, I exclude a price of one-half where herding is precluded. All four results are highly significant ($p = 0.000$).

Contrarian behavior also increases in each treatment as predicted, but insignificantly so (MAIN, $p = 0.204$; NO BELIEFS, $p = 0.376$), perhaps reflecting the fact that contrarian behavior is not frequent enough to provide sufficient statistical power. Finally, abstention as well, but it is not significant ($p = 0.305$). Results when partial contrarian trades are included are similar.
marginally significantly decreases in the NO BELIEFS treatment ($p = 0.047$) as predicted, but increases in the MAIN treatment ($p = 0.005$) because of the tendency of some subjects to abstain at extreme prices. Standard expected utility with a large degree of risk aversion predicts this behavior, and in fact I show in the following section that a handful of subjects are best described by these preferences.

**Result 3:** Revealing behavior (always trading according to one’s private signal) decreases with the normalized price and herding increases, as predicted by the model. Contrarian behavior increases as predicted, but not significantly so. Abstention decreases as predicted in the NO BELIEFS treatment, but increases in the MAIN treatment.

At the individual level, the model predicts that an individual makes herding or contrarian decisions, but not both. To test this prediction, Figure 4 plots the fraction of contrarian decisions versus the fraction of herding decisions for each subject in each treatment. If the theory were perfect, we’d expect ‘contrarian’ individuals to lie on the y-axis and ‘herding’ individuals on the x-axis. Although noisy, there is a clear negative correlation in both treatments, indicating that most subjects have a clear predisposition for either herding or contrarianism. This predisposition allows me to classify subjects according to their preferences (herding, contrarian, or expected utility) as illustrated in the figure and described in detail in the following section.

**Result 4:** The majority of subjects have a predisposition towards either herding or contrarianism, consistent with the model.

### 3.2.2 Preference Estimation

The previous section provides evidence consistent with CPT preferences, but does CPT better describe decisions than expected utility? I begin with a theoretical result that suggests expected utility is unlikely to fit the data well because of a limited ability to generate unresponsive trading and abstention, both of which are common in the data.

**Lemma 2:** Consider an investor with standard expected utility preferences:

(i) if risk-averse (strictly concave utility function), she does not exhibit herding or contrarian behaviors.

(ii) if risk-seeking (strictly convex utility function), she does not abstain for either private signal.

The intuition behind Lemma 2 is straightforward. For part (i), herding and contrarianism
involve taking a negative expected value bet for one of the private signal realizations, so that a risk-averse subject instead prefers to abstain. For part (ii), given that a risk-neutral subject prefers the positive expected value bet of trading with her signal over abstaining, a risk-seeking subject does as well, and so will not abstain.

In light of Lemma 2, it is still possible that the aggregate data could be generated by a mixture of risk-averse and risk-seeking types. To determine whether or not this is the case, I compare the fit of expected utility to that of prospect theory. I focus on the MAIN treatment where, because Bayesian errors play no role, behavior provides the cleanest expression of preferences. I begin by imposing the same general model on all subjects in the treatment, but allowing the preference parameters to vary on an individual basis.

I consider four candidate preference models. The first is CPT. As I show in Appendix B, it has two degrees of freedom: one relating to loss aversion, $\lambda$, and the other relating to the difference between the utility curvature and probability weighting parameters, $\delta - \alpha$. In the model, I constrain $\alpha$ and $\delta$ to lie in the interval, $(0, 1]$, which generates restrictions on the trade patterns CPT can generate (see Appendix B for details).27 The second is CPT preferences without loss aversion ($\lambda = 1$), giving it a single degree of freedom and making it directly comparable to the third model, expected utility. For expected utility, in the reported results, I assume CRRA preferences, but the results are almost identical with CARA preferences. Finally, I consider a non-parametric model which only requires that

---

27In making this restriction, I'm imposing an S-shaped utility function and an inverse S-shaped probability weighting function consistent with the findings of Kahneman and Tversky (1979). The restriction could be relaxed to improve the fit of CPT.
decisions respect symmetry (the decision at a price \( p_t \) with a favorable signal must match that at a price \( 1 - p_t \) with an unfavorable signal). I can construct this ‘best symmetric’ model because of the discrete nature of the beliefs at which subjects make decisions. It has 27 degrees of freedom and encompasses expected utility, CPT, and any other model that respects symmetry as special cases, thus providing an upper bound on the ability of any such model to fit the data (see Appendix C for further discussion).

To compare models, I employ the technique of Bisiere, Decamps, and Lovo (2015) to calculate a match score for each subject relative to each model of behavior. For a given model and a given (range of) preference parameters, I obtain predicted decisions for each discrete price and for each private signal. I award 0.5 for each of a subject’s 54 decisions that matches the prediction (and zero otherwise), and then divide by the maximum possible score so that the match score lies between 0 and 1.

This technique of constructing a match score is admittedly simplistic, but, in this specific setting, a more sophisticated method of preference estimation using maximum likelihood is likely to be misleading. Upon writing down a standard random utility model, one sees immediately that, allowing for stochastic errors causes \( \alpha \) and \( \delta \) to enter the likelihood function independently. Therefore, it would seem to be possible to estimate both \( \alpha \) and \( \delta \), even though in the absence of errors, I’ve established that only their difference matters (see Appendix B). In this case, the estimates of \( \alpha \) and \( \delta \) obtained would depend critically on the distribution of errors assumed (see also Wilcox (2008)). Rather than obtain potentially misleading point estimates of both \( \alpha \) and \( \delta \), I instead opt for a well-identified range of their difference.

Figure 5 provides the empirical CDFs of the match scores for each of the four models. A particularly striking result is that CPT explains choices as well as the best possible model that imposes symmetry, even though it has an order of magnitude less degrees of freedom (\( p = 0.269 \) with a Kolmogorov-Smirnov test). Furthermore, even restricting prospect theory to not include loss aversion does almost as well as the full prospect theory model, and does significantly better than expected utility, both of which have a single degree of freedom (\( p = 0.008 \) in a Kolmogorov-Smirnov test).

Rather than impose a single class of preferences on all subjects, we can allow it to vary by subject. Specifically, for each subject, I find the parameters which provide the best fit

---

28This problem of ‘incredible structural inference’ is nicely illustrated by Thomas Rosenberg (2005) with the example of a research assistant that repeatedly measures the length and width of a table, but reports only the table’s area. Absent measurement error, it is clear that one cannot recover the length and width from the area. However, with assumptions on distributions of the measurement errors in each dimension, one can miraculously recover the two dimensions.

29The best model that imposes symmetry matches about 75% of a subject’s decisions at the median, similar to the finding of Bisiere, Decamps, and Lovo (2015). Buying is more frequent in both treatments, perhaps because it is more familiar to subjects than selling short.
Figure 5: Model Fits in MAIN Treatment

Note: Empirical CDFs of the match scores of subjects for expected utility, prospect theory without loss aversion, prospect theory with loss aversion (thick line), and the best possible model that imposes symmetry.

(give the highest match score) for both CRRA and CPT preferences. I then take the better of the two, breaking ties in favor of expected utility. Table 3 provides the number of subjects for which each model provides the best match, along with their preference type: risk-neutral ($\alpha = \delta = 1$), risk-averse (CRRA with $\alpha < 1$), risk-seeking (CRRA with $\alpha > 1$), herding ($\alpha < \delta$), or contrarian ($\alpha > \delta$). The first row restricts CPT to a single degree of freedom (no loss aversion) and the second considers the full model.

Table 3 shows that 76% of subjects are better described by prospect theory, which is interestingly similar to the 80% of subjects that Bruhin, Fehr-Duda, and Epper (2010) report.

<table>
<thead>
<tr>
<th>Table 3: Individual Preference Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>Both</td>
</tr>
<tr>
<td>Both</td>
</tr>
</tbody>
</table>

Note: Number of subjects best matched to each model of behavior. In the first row, CPT is restricted to not include loss aversion. In the second, full CPT is considered.
Remarkably, there is little difference in the classification of subjects when loss aversion is included. Within the set of subjects for which CPT is the best match, approximately 90% exhibit herding behavior. Importantly, however, contrarian types exist, which is consistent with CPT but inconsistent with the other non-expected utility theory that has been studied in this environment, regret aversion (Qin (2015)). Among the subjects for which expected utility is a better match, the majority are classified as risk-neutral which is technically consistent with both CPT and expected utility. However, three subjects are classified as risk-averse, exhibiting behavior that is inconsistent with CPT. These subjects abstain as prices approach zero or one, consistent with a high level of standard risk-aversion. Returning to the graph for the MAIN treatments in Figure 4, we see that subjects classified as herding or contrarian types have a clear disposition for their preferred behavior, with many never making a single decision that contradicts their ‘type’.

**Result 5:** Approximately three-quarters of subjects in the MAIN treatment are better described by prospect theory than expected utility (CRRA) and, for these, herding strategies are dominant. Furthermore, prospect theory fits the data as well as any model that imposes symmetry.

Figure 6 plots the estimates of the parameters for the 43 subjects that have preferences consistent with CPT (thus including risk-neutral subjects). Although the sign of $\delta - \alpha$ fully determines whether a subject is a herding or contrarian type, its magnitude is not uniquely determined (see Appendix B for details). I normalize $\delta = 1$ for herding types and $\alpha = 1$ for contrarian types in order to keep the parameters as close to the risk-neutral benchmark ($\delta = \alpha = 1$) as possible, while ensuring both parameters remain weakly less than one. From the range of parameters consistent with an individual, I report the smallest possible $\lambda$ and difference in $\delta - \alpha$.

In Figure 6, we observe considerable heterogeneity: 4 contrarian subjects have $\delta - \alpha < 0$, 7 risk-neutral subjects have $\delta - \alpha = 0$, and the remaining 32 herding type subjects have $\delta - \alpha > 0$. The median and modal value of $\delta - \alpha$ is 0.5, corresponding to herding at prices, $p_t \geq 0.84$ and $p_t \leq 0.16$. The original finding of Kahneman and Tversky (1992) that $\delta < \alpha$ has led to a belief that probability weighting dominates utility curvature, in contrast to my findings. However, a closer examination of the literature shows that parameter estimates vary widely and my finding is nothing out of the ordinary: across the ten preference measurement studies summarized in Table A.3 of Glimcher and Fehr (2013) that use the same probability weighting function I use, four also find $\delta > \alpha$ at the median. Most of my estimates for $\lambda$ lie between one and two, consistent with a large literature. Unfortunately for comparison purposes, very little evidence on the overall joint distribution exists. Only a single paper I’m
Figure 6: Parameter Estimates in MAIN Treatment

Note: The size of a dot indicates the number of subjects with those preference estimates, ranging from 1 (smallest) to 12 (largest).

I aware of (Zeisberger, Vrecko, and Langer (2012)) reports individual preference parameters. They document substantial heterogeneity as I do, but, importantly, also that estimates for the majority of subjects are stable across elicitation tasks one month apart.

3.2.3 Bayesian Errors

Assuming preferences are drawn from the same population in the MAIN and NO BELIEFS treatments, we can use the estimated preferences from the MAIN treatment to assess the types of Bayesian errors that lead to differences in behavior across treatments. Specifically, for the given estimated distribution of preference parameters and a given belief, I calculate the average frequency of buy, sell, and no trade decisions in the population. I vary the belief until I find the belief closest to the correct Bayesian belief (in case of a tie) which best matches the empirical frequencies of buy, sell, and no trade decisions in the NO BELIEFS treatment (in terms of minimizing the sum of squared errors). I repeat this exercise for each price and each private signal. Figure 7 plots the estimated beliefs with correct Bayesian beliefs as the benchmark.

Figure 7 suggests that subjects tend to underweight their private information both for favorable and unfavorable signals at intermediate public beliefs/prices: the distance between

30 In this exercise, I assume a population of risk-neutral and CPT types, dropping the 3 risk-averse subjects.
their beliefs for favorable and unfavorable signals is too narrow. This result closely tracks the finding of Bisiere, Decamps, and Lovo (2015), and is consistent with the fact that there is more abstention at intermediate public beliefs in the NO BELIEFS treatment (Table 2). At more extreme prices, the information contained in the price is underweighted as beliefs are not extreme enough, generating the increase in partial contrarian and contrarian behaviors in the NO BELIEFS treatment: the asset is perceived to be too expensive at high prices and too cheap at low prices. Bisiere, Decamps, and Lovo (2015) find exactly the opposite result. In their data, beliefs that are too extreme are required to reconcile the herding that is present in their data (which can’t be explained by the expected utility preferences they consider). Here, instead, subjects have a natural tendency for herding which is counteracted by underweighting the information contained in the price. This underweighting of public signals (which generate the price) is consistent with the conclusion of Benjamin (2018) that, across many studies, underinference from information is the norm.

4 Informational Efficiency

In this section, I simulate the theoretical model to determine the extent to which unresponsive behavior affects the informational efficiency of market prices. To do so, I consider an
extension of the model that allows for multiple CPT preference types (the full estimated distribution). As in the original model, I assume common knowledge of the distribution of preferences so that prices reflect the equilibrium informational content of trades. I simulate the model first without Bayesian errors (reflecting behavior in the MAIN treatment) and then allowing for Bayesian errors (reflecting behavior in the NO BELIEFS treatment).  

Informational efficiency can be measured by the distance of the price (the expected value of the asset) from the true asset value, which I fix \((V = 1)\) in the simulations.\(^{32}\) Starting from a uniform prior \((p_1 = \frac{1}{2})\), we know that because of risk-neutral investors, prices will eventually converge to the true asset value (Section 2.3), but how rapidly? Figure 8 plots quartiles of the simulated price across time. As a benchmark, I simulate price convergence assuming all investors are risk-neutral. In all simulations, I use \(q = 0.7\) as in the experiment.

Figure 8 illustrates the reductions in informational efficiency caused by CPT preferences: prices take longer to converge at all quartiles of the price distribution. Although moderate in absolute terms, in relative terms the differences are substantial. For example, absent Bayesian errors, the median distance between the price and the true asset value is negligible

\(^{31}\)As in the previous section, I continue to exclude the three risk-averse subjects in this exercise. To construct non-Bayesian beliefs at public beliefs that are not observed in the data, I linearly interpolate between the nearest estimated beliefs.

\(^{32}\)The standard measure of pricing error is \(\text{Var}(V|H_t) = p_t(1 - p_t)\), but it seems more natural to use a simple distance metric when the asset value is known, as in the simulations. Convergence of the bid and ask prices tracks convergence of the expected value closely so I focus on the expected value for simplicity.
After twenty periods. But, under the estimated distribution of preferences, it increases by roughly a factor of 50 to 0.05.

The main reason for slower convergence is the large proportion of herding trades generated by prospect theory preferences. For prices greater than one-half, herding trades slow price convergence to the true asset value because they reduce the amount of information revealed by making some purchases uninformative (leaving the information revealed by a sale unaffected). This intuition reverses for prices less than one-half: price convergence is sped up because sales become less informative ensuring incorrect signal realizations have less impact. On average, because prices are more likely to be above one-half when \( V = 1 \), herding slows price convergence.

Comparing the two graphs in Figure 8, we see that Bayesian errors do not substantially affect the median pricing error, but they do widen the range of pricing errors. Price convergence is faster in the top quartile, but much slower in the bottom quartile. To understand why, note that when Bayesian errors are introduced, herding and partial herding are replaced by contrarian and partial contrarian behaviors (see Table 2). Contrarianism has the opposite effect on price convergence that herding behavior does - it speeds convergence at prices greater than one half by making sales less informative and slows convergence at prices less than one half by making purchases less informative. Thus, when herding trades are replaced by contrarian trades, if initial signal realizations are correct (top quartile), convergence speeds up. However, if initial signal realizations are incorrect (bottom quartile), convergence slows. These results indicate that it is not only the presence of unresponsive behavior that affects informational efficiency, but also it’s composition.

5 An Extension to Actual Market Returns

In the model, the relationship between price trends and the skewness in returns is driven by the binary nature of the asset value, limiting the model’s application to binary assets or ‘near binary’ assets such as options and initial public offerings (Green and Hwang (2012)). However, previous research has noted the same relationship between rising prices and negatively skewed returns (and conversely) is present in actual market data (Harvey and Siddique (1999) and Chen, Hong, and Stein (2001)). Does the intuition of the model - that skewness in returns drives herding - carry over to the very complex return distributions of actual markets? To answer this question, I simulate the behavior of the modal subject from the experiment when facing these returns.

I first demonstrate the relationship between price trends and skewness in the daily NYSE/AMEX/NASDAQ index return data from CRSP (value-weighted, including distri-
Table 4: Conditional Moments of the Market Index Daily Returns

<table>
<thead>
<tr>
<th>Price Level</th>
<th>Moment</th>
<th>Estimate</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Price</td>
<td>Mean</td>
<td>0.00041</td>
<td>(-0.00027, 0.0011)</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.018</td>
<td>(0.017, 0.019)</td>
</tr>
<tr>
<td></td>
<td>Skew</td>
<td>0.65</td>
<td>(0.18, 1.22)</td>
</tr>
<tr>
<td>High Price</td>
<td>Mean</td>
<td>0.0064</td>
<td>(0.00022, 0.0011)</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.011</td>
<td>(0.010, 0.012)</td>
</tr>
<tr>
<td></td>
<td>Skew</td>
<td>-0.52</td>
<td>(-1.21, 0.20)</td>
</tr>
</tbody>
</table>

Note: Mean, standard deviation, and skewness of the daily returns from two constructed conditional distributions. The High Price distribution comes from days in the month following a six-month return in the highest decile. The Low Price distribution comes from days in the month following a six-month return in the lowest decile.

Table 4 provides these measures for months in which the previous six-month returns were in the highest decile of past returns (‘High Price’) and the lowest decile (‘Low price’), including the corresponding 95% confidence intervals calculated using a percentile bootstrap. The results show that the index returns become positively skewed after price increases and negatively skewed after price decreases, with little difference in mean returns.

If skewness drives herding, the results in Table 4 suggest that we should expect CPT investors to buy after rising prices and sell after falling prices when faced with actual market returns. To test this conjecture, I consider the trading decision of the modal subject from the MAIN treatment (α = 0.5 and δ = λ = 1) who faces the return distribution associated with either a high or low price level. For simplicity, I model the investor’s private information as information about the mean expected return, and calculate the private signal that would make her indifferent between buying and selling in each case. Table 5 reports the corresponding expected returns conditional on this private signal (annualized), along with their 95% confidence intervals. For comparison purposes, note that these private expected

---

33The results are very similar if October 1987 is excluded, ensuring that they are not driven by the market crash during this period.

34Skewness is calculated with $SKEW = \frac{E(r - \overline{r})^3}{STD(r)^3}$ where $\overline{r}$ is the mean return and $STD(r)$ is the standard deviation of returns.

35Because $\lambda = 1$ and the decision weight functions are the same for losses and gains, the utility from buying is equal to the opposite of the utility from selling. Therefore, indifference between buying and selling also implies indifference with abstention.
Table 5: Annualized Private Returns for CPT Investor

<table>
<thead>
<tr>
<th>Price Level</th>
<th>Estimate</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Price</td>
<td>7.1%</td>
<td>(0.0%, 15.7%)</td>
</tr>
<tr>
<td>High Price</td>
<td>−7.3%</td>
<td>(−11.6%, −2.8%)</td>
</tr>
</tbody>
</table>

Note: Each return is the annualized return for which a CPT investor (α = 0.5, λ = δ = 1) is indifferent between buying and selling. It is calculated assuming the investor faces the distribution of historic daily returns conditional on the return over the previous six months being in the upper or lower decile of such returns.

returns would be zero for a risk-neutral investor.\textsuperscript{36}

Table 5 shows that, at low prices, a CPT investor prefers to sell the asset (herd) if her expected return, conditional on private information, is 7.1% or less. At high prices, she instead prefers to buy the asset (herd) if her private expected return is −7.3% or greater. Thus, a CPT investor requires substantial returns to trade against prior trends. Although the confidence intervals for these estimates are wide, they do not overlap, establishing a significant difference. The fact that CPT preferences can generate herding in the presence of very complex return distributions demonstrates that the results of this paper are not trivially related to the binary environments considered, but are instead much more generally applicable.

6 Discussion

Herding and contrarian strategies are of interest because of their potential for reducing the informational efficiency of prices by not revealing information to the market. These strategies are often thought to rely on (or even be defined as relying on) the observation of past histories of actions and/or prices. In this paper, however, I have shown both theoretically and via a laboratory experiment that they can emerge from preferences. The strategies depend only upon past price trends to the extent that these trends are related to future returns. In the experiment, I control for social and belief-based causes of behavior, yet observe economically significant levels of herding and contrarianism, demonstrating a clear role for preferences. Among the three-quarters of subjects whose behavior is more consistent with prospect theory

\textsuperscript{36}For example, if the market return is −5%, a risk-neutral investor would require a private signal of 5% to make her indifferent. The private returns she faces is then −5 + 5 = 0%.
Table 6: NO BELIEFS versus Cipriani and Guarino (2009)

<table>
<thead>
<tr>
<th>Normalized Price</th>
<th>Herding</th>
<th>Contrarian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO BELIEFS</td>
<td>CG2009</td>
</tr>
<tr>
<td>0.50</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.70</td>
<td>10.0</td>
<td>5.7</td>
</tr>
<tr>
<td>0.84</td>
<td>13.9</td>
<td>16.1</td>
</tr>
<tr>
<td>0.93</td>
<td>23.1</td>
<td>23.9</td>
</tr>
<tr>
<td>≥0.97</td>
<td>21.7</td>
<td>21.9</td>
</tr>
</tbody>
</table>

Note: Percentages of herding and contrarian decisions in the NO BELIEFS treatment and from Table 2 in Cipriani and Guarino (2009), CG2009.

than expected utility, almost 90% of subjects exhibit a preference for herding strategies.\(^{37}\)

This high proportion of herding is consistent with a long-standing narrative of financial markets - investors chase price trends, producing bubbles and crashes (Mackay (1841)). In the context of previous experimental findings, however, the results are surprising. Cipriani and Guarino (2005,2009) and Drehmann, Oechssler, and Roider (2005) each report more contrarian than herding behavior. How to reconcile these findings and to what extent may prospect theory preferences be responsible for behavior in past experiments? To answer these questions, consider the experiment of Cipriani and Guarino (2009), which is notable because they conduct their experiment with a sample of financial market professionals so that the extent to which my data corresponds to theirs provides some assurance that my results are not particular to an undergraduate student population.\(^{38}\) Table 6 compares the frequencies of herding and contrarian decisions in my NO BELIEFS treatment (subjects in Cipriani and Guarino (2009) had to perform Bayesian updating) with those reported in Cipriani and Guarino (2009), Table 2.

In Table 6, we see that the frequencies of herding at each price are remarkably similar across the two experiments, and are not insignificant, exceeding 20% of decisions as prices become extreme. This finding suggests that prospect theory is responsible for a significant fraction of decisions in Cipriani and Guarino (2009) because no other available theory generates herding in this environment. On the other hand, the level of contrarianism is much higher in Cipriani and Guarino (2009), which leads to an intriguing possibility. In their setup, subjects had to form beliefs about the strategies of other subjects. If subjects understand

\(^{37}\)In a paper that motivated this project (Kendall, 2018), I also observe frequent herding. There, because the environment is much more complex (subjects choose not only the direction of trade, but when to trade), it is difficult to draw strong conclusions about the mechanism driving this behavior.

\(^{38}\)Cipriani and Guarino (2009), like I, use the strategy method and set \(q = 0.7\) which also makes the two studies more comparable.
that some fraction of subjects herd (an understanding that experienced financial professionals may be more likely to have), it can be a best response to act contrarian, providing an additional reason to act contrarian in their experiment that is not present in mine. Of course, the differences across experiments may also simply reflect differences in the preferences of the two populations. Understanding the extent to which subjects form and respond to correct beliefs about the preferences of others in this (and related) environments is an interesting question for future research.

Interpreted as a preference for negative skewness, my results are consistent with those of Huber, Palan, and Zeisberger (2017). They document that positively-skewed assets are more likely to be perceived as risky (having a higher probability of a loss) and thus trade at lower prices than negatively-skewed assets. In real-world markets, the evidence has tended to point to preferences for positive skewness (Kumar (2009), Green and Hwang (2012), Eraker and Ready (2015)). Unlike the laboratory evidence, however, this evidence reflects selection into these markets (as documented by Kumar (2009)).

Prospect theory is, at its core, about framing - people perceive gambles differently when framed as gains or losses (Kahneman and Tversky (1979,1981)). Financial markets frame choices in a particular way, one which I adopted in the experimental design. But, the potential for framing effects raises the intriguing possibility that the results of the experiment would be different if framed in some other way. In fact, Bisiere, Decamps, and Lovo (2015) provide some evidence for a framing effect: they find more herding when decisions are framed as a market in their robustness treatment (SME) than as a choice between lotteries (LE).
If framing matters, the way in which market data is presented may affect the extent of herding and thus the informational efficiency of the market. Exploring this possibility is an interesting, and potentially very important, topic for future research.

A broader theme suggested by this paper is that prospect theory preferences can play a critical role in market and game experiments. In interpreting the data from these types of experiments, risk-neutral (expected utility) preferences are often assumed, citing the Rabin critique (Rabin (2000)): if people are expected utility maximizers, then they must be risk-neutral in the laboratory. But, Rabin’s conclusion is not that we should assume risk-neutrality. Instead, he argues that because we often see departures from risk-neutrality over small stakes, we should consider models other than expected utility. Failing to do so can lead to very erroneous conclusions about the underlying reasons for any observed departures from ‘rational’ behavior.

References


[63] Weizsäcker, G. (2010), ”Do we follow others when we should? A simple test of rational expectations”, *American Economic Review*, 100 (5), 2340-2360


### Appendices

#### A. Omitted Proofs

*Proof of Lemma 1:*

For convenience, I refer to investors with favorable signals \((s_t = 1)\) as type 1, and investors with unfavorable signals \((s_t = 0)\) as type 0. I provide the proof that a type 1 risk-neutral investor buys. The proof that a type 0 risk-neutral investor sells is symmetric.

I first claim that if a type 1 informed investor (risk-neutral or CPT) sells at some \(p_t\) with positive probability in equilibrium, then a type 0 investor must also sell at \(p_t\) with probability one. For the CPT investor, this fact follows from inspecting equations (5) and their counterparts for a type 0 investor. For risk-neutral investors, if a type 1 investor sells with positive probability, then it must be that case that \(B_t - b_1^t \geq b_1^t - A_t\) so that she weakly prefers selling to buying. Rearranging, \(B_t + A_t \geq 2b_1^t > 2b_0^t\) so that the type 0 investor must strictly prefer selling. Similarly, if a type 1 investor weakly prefers selling to not trading, the type 0 investor must strictly prefer the same.

Given that type 0 investors must sell if type 1 investors sell, it follows that a sell trade either reveals no information or negative information. Therefore, the bid price must be weakly less than the public belief, \(B_t \leq p_t\) which implies that a type 1 risk-neutral investor will never sell
because her expected profit is negative: \( B_t - b_t^1 < B_t - p_t \leq 0 \) (using \( b_t^1 > p_t \)). I now show that she also never abstains.

Using the formula for the ask price, (4), a type 1 risk-neutral investor prefers buying over abstaining if

\[
\frac{b_t^1 - A_t}{p_t q (1 - p_t) (1 - q)} > 0 \quad \text{so that}\quad \frac{p_t Pr(a_t = \text{buy}|V = 1)}{Pr(a_t = \text{buy}|V = 1) + (1 - p_t) Pr(a_t = \text{buy}|V = 0)} > \frac{q}{1 - q} 
\]

where \( Pr(a_t = \text{buy}|V = 1) \) and \( Pr(a_t = \text{buy}|V = 0) \) depend upon the equilibrium strategies of the informed investors,

\[
Pr(a_t = \text{buy}|V = 1) = \frac{1 - \mu}{2} + \mu \gamma \beta^\text{RN}(V = 1) + \mu(1 - \gamma)\beta^\text{PT}(V = 1) \\
Pr(a_t = \text{buy}|V = 0) = \frac{1 - \mu}{2} + \mu \gamma \beta^\text{RN}(V = 0) + \mu(1 - \gamma)\beta^\text{PT}(V = 0) 
\]

and \( \beta^\text{RN}(V = x) \) and \( \beta^\text{PT}(V = x) \), \( x \in \{0, 1\} \) are the probabilities of observing buy orders from risk-neutral and CPT investors, conditional on \( V = x \), respectively. \( \beta^\text{PT}(V = 1) = q \beta^\text{1,PT} + (1 - q) \beta^\text{0,PT} \) and \( \beta^\text{PT}(V = 0) = (1 - q) \beta^\text{1,PT} + q \beta^\text{0,PT} \) where \( \beta^\text{y,PT} \) is the probability that the market maker believes a CPT investor with \( s_t = y \) buys. The right-hand side of equation (7) can be shown to be strictly increasing in \( \beta^\text{1,PT} \) and strictly decreasing in \( \beta^\text{0,PT} \) so that it can be bounded above by the case of \( \beta^\text{1,PT} = 1 \) and \( \beta^\text{0,PT} = 0 \).

Furthermore, if a type 1 risk-neutral investor were to abstain, then a type 0 risk-neutral investor must also abstain (this fact is established in the same manner as the fact that, if a type 1 risk-neutral investor sells, then a type 0 risk-neutral investor must also sell), so that \( \beta^\text{RN}(V = 1) = \beta^\text{RN}(V = 0) = 0 \). Thus, if a type 1 risk-neutral investor were to abstain, we must have

\[
q \leq \frac{1 - \mu}{2} + \mu(1 - \gamma)q \\

\text{But, this inequality never holds for } \mu < 1. \text{ Conversely, a type 1 risk-neutral investor strictly prefers buying to abstaining (so that } \beta^\text{RN}(V = 1) = q \text{ and } \beta^\text{RN}(V = 0) = 1 - q \text{) because a type 0 risk-neutral investor can never buy, just as a type 1 risk-neutral investor can never sell) if}
\]

\[
q \geq \frac{1 - \mu}{2} + \mu \gamma q + \mu(1 - \gamma)q \\

\text{This inequality holds for all } \mu < 1. \square
\]

**Proof of Theorem 1:**
Part 1 follows directly from the assumption that the market maker faces perfect competition and Bayes’ rule.
Part 2 is proven in Lemma 1.
Part 3a. Given \( \alpha = \delta \), the optimality conditions for a type 1 CPT investor, (5), become

\[
\begin{align*}
\text{buy if } 1 &\geq \lambda \left( \frac{1 - q}{q} \frac{Pr(a_t = \text{buy}|V = 1)}{Pr(a_t = \text{buy}|V = 0)} \right)^\alpha \\
\text{sell if } 1 &\leq \frac{1}{\lambda} \left( \frac{1 - q}{q} \frac{Pr(a_t = \text{sell}|V = 1)}{Pr(a_t = \text{sell}|V = 0)} \right)^\alpha
\end{align*}
\]

From (9), and the corresponding equations for a type 0 investor (in which the ratios of \( q \)
and $1-q$ are inverted), we see that whether or not an investor trades is independent of the current public belief.

I first show that a type 0 investor can never buy with positive probability. Using the equilibrium strategies of the risk-neutral investor, we can write the probabilities of observing a buy conditional on $V = 1$ and $V = 0$ as

$$
Pr(a_t = buy|V = 1) = \frac{1-\mu}{2} + \mu \gamma q + \mu (1-\gamma) \beta^{PT}(V = 1)
$$

$$
Pr(a_t = buy|V = 0) = \frac{1-\mu}{2} + \mu \gamma (1-q) + \mu (1-\gamma) \beta^{PT}(V = 0)
$$

where, as in the proof of Lemma 1, $\beta^{PT}(V = 1) = q \beta^{1,PT} + (1-q) \beta^{0,PT}$ and $\beta^{PT}(V = 0) = (1-q) \beta^{1,PT} + q \beta^{0,PT}$. Now, as argued in the proof of Lemma 1, if a type 0 investor buys, then so must a type 1 investor. This fact implies $\beta^{1,PT} \geq \beta^{0,PT}$ which in turn implies that the ratio $\frac{Pr(a_t = buy|V = 1)}{Pr(a_t = buy|V = 0)}$ is bounded below by $\beta^{1,PT} = \beta^{0,PT}$, because this ratio is increasing in $\beta^{1,PT}$ and decreasing in $\beta^{0,PT}$. Therefore, for a type 0 investor to buy, we must have

$$
1 \geq \lambda \left( \frac{1-q}{q} \frac{1-\mu}{2} + \mu \gamma q + \mu (1-\gamma) q \beta^{1,PT} \right)^\alpha.
$$

But, the ratio inside the the parentheses is strictly greater than one for any $\beta^{0,PT}$, so does not hold for any $\lambda \geq 1$.

Given that a type 0 investor never buys and, as can be shown similarly, a type 1 investor never sells, we are left to determine the conditions under which investors trade according to their private information, and when they do not trade. Substituting the probabilities of observing a buy into the first equation of (9) (using the fact that a type 0 investor never buys ($\beta^{0,PT} = 0$)), we have that a type 1 investor buys if

$$
\text{buy if } 1 \geq \lambda \left( \frac{1-q}{q} \frac{1-\mu}{2} + \mu \gamma q + \mu (1-\gamma) q \beta^{1,PT} \right)^\alpha.
$$

For $\lambda$ sufficiently large, the investor will not buy. Setting $\beta^{1,PT} = 0$, we can find the cutoff value of $\lambda$, $\overline{\lambda}$

$$
1 \leq \lambda \left( \frac{1-q}{q} \frac{1-\mu}{2} + \mu \gamma q \right)^\alpha
$$

$$
\iff \lambda \geq \left( \frac{1-q}{q} \frac{1-\mu}{2} + \mu \gamma q \right)^\alpha \equiv \overline{\lambda}.
$$

For $\lambda$ sufficiently small, the investor will buy with probability one. Setting $\beta^{1,PT} = 1$, we can find the cutoff value of $\lambda$, $\underline{\lambda}$

$$
1 \geq \lambda \left( \frac{1-q}{q} \frac{1-\mu}{2} + \mu \gamma q \right)^\alpha
$$

$$
\iff \lambda \leq \left( \frac{1-q}{q} \frac{1-\mu}{2} + \mu \gamma q \right)^\alpha \equiv \underline{\lambda}.
$$

Simple algebra shows that $\overline{\lambda} > \underline{\lambda} > 1$ for all parameterizations. Finally, for intermediate values of $\lambda$, the investor mixes between buying and not trading such that the ask price makes him indifferent between the two. The mixing probability, $\beta^{1,PT}$, satisfies

$$
1 = \lambda \left( \frac{1-q}{q} \frac{1-\mu}{2} + \mu \gamma q + \mu (1-\gamma) q \beta^{1,PT} \right)^\alpha
$$

which has a unique solution for $\beta^{1,PT}$. The conditions for a type 0 investor to always sell, not trade, and mix between not trading and selling are easily shown to be identical to those for a type 1 investor.

Part3b. Consider $\delta > \alpha$. I first evaluate the decisions to buy for both types. Beginning with
equation (5) and the formulae for the probability of observing a buy (8), the two equations governing buy decisions are given by

\begin{equation}
\begin{align*}
s_t = 1 : \left( \frac{p_t}{1-p_t} \right)^{\delta-\alpha} & \geq \lambda \left( \frac{1-q}{q} \right)^{\delta} \left( \frac{\frac{1-\mu}{2} + \mu \gamma q + \mu (1-q) q^{\beta_1,PT} + (1-q) \beta_0,PT}{\frac{1-\mu}{2} + \mu \gamma q + \mu (1-q) q^{\beta_1,PT} + q^{\beta_0,PT}} \right)^{\alpha} \\
s_t = 0 : \left( \frac{p_t}{1-p_t} \right)^{\delta-\alpha} & \geq \lambda \left( \frac{q}{1-q} \right)^{\delta} \left( \frac{\frac{1-\mu}{2} + \mu \gamma q + (1-q) q^{\beta_1,PT} + (1-q) \beta_0,PT}{\frac{1-\mu}{2} + \mu \gamma q + (1-q) q^{\beta_1,PT} + q^{\beta_0,PT}} \right)^{\alpha}
\end{align*}
\end{equation}

(10)

Because the left-hand side of each equation is an unbounded function of the public belief and the right-hand side of each is independent of the public belief, we immediately see that both types must buy for sufficiently high public beliefs. Furthermore, because the ask price always exceeds the bid price, from (3), we can see that each type of CPT investor must not trade for some range of public beliefs.\(^{43}\) It remains to characterize behavior in the transition region from not trading to buying.

From the proof of Lemma 1, we know the type 1 investor must buy with probability one if the type 0 investor buys with positive probability, so that in the transition, \(p^1\), of the type 1 investor, the type 0 investor must buy with probability zero. In this case, the type 1 investor buys if

\begin{equation}
\left( \frac{p_t}{1-p_t} \right)^{\delta-\alpha} \geq \lambda \left( \frac{1-q}{q} \right)^{\delta} \left( \frac{\frac{1-\mu}{2} + \mu \gamma q + \mu (1-q) q^{\beta_1,PT}}{\frac{1-\mu}{2} + \mu \gamma (1-q) + \mu (1-q) (1-q) q^{\beta_1,PT}} \right)^{\alpha}
\end{equation}

(11)

The right-hand side of (11) is monotonically increasing in \(q^{\beta_1,PT}\), which ensures a unique fixed point between the market maker’s belief and the action of the type 1, CPT investor. In particular, as the public belief increases, \(q^{\beta_1,PT}\) must increase such that the equation holds with equality: the type 1 investor mixes with \(q^{\beta_1,PT} \in (0, 1)\). Intuitively, the more that the type 1 investor buys, the more information revealed by her purchase, the higher the ask price, and the less profitable it is to buy. The lower and upper bounds of the transition, \(p^1\) and \(\bar{p}^1\), satisfy

\begin{equation}
\begin{align*}
\left( \frac{p^1}{1-p^1} \right)^{\delta-\alpha} &= \lambda \left( \frac{1-q}{q} \right)^{\delta} \left( \frac{\frac{1-\mu}{2} + \mu \gamma q}{\frac{1-\mu}{2} + \mu \gamma (1-q)} \right)^{\alpha} \\
\left( \frac{\bar{p}^1}{1-\bar{p}^1} \right)^{\delta-\alpha} &= \lambda \left( \frac{1-q}{q} \right)^{\delta} \left( \frac{\frac{1-\mu}{2} + q}{\frac{1-\mu}{2} + \mu (1-q)} \right)^{\alpha}
\end{align*}
\end{equation}

(12)

In the transition region of the type 0 investor, \(1-p^0\), the type 1 investor buys with probability one so that the type 0 investor buys if

\begin{equation}
\left( \frac{p_t}{1-p_t} \right)^{\delta-\alpha} \geq \lambda \left( \frac{q}{1-q} \right)^{\delta} \left( \frac{\frac{1-\mu}{2} + \mu \gamma q + \mu (1-q) q^{\beta_0,PT} + (1-q) \beta_0,PT}{\frac{1-\mu}{2} + \mu \gamma (1-q) + \mu (1-q) (1-q) q^{\beta_0,PT} + q^{\beta_0,PT}} \right)^{\alpha}
\end{equation}

(13)

Contrary to the case of the type 1 investor, the right-hand side of (13) is monotonically decreasing in \(q^{\beta_0,PT}\). Thus, for a given public belief, multiple fixed points between the market maker’s belief and the type 0 CPT investor’s action may exist. However, Bertrand compe-

\(^{43}\)The ask price always exceeds the bid price because risk-neutral investors always trade according to their private information and CPT investors’ trades either reveal their private information or no information. We cannot have only a type 0 investor buying, for example, because the proof of Lemma 1 shows that the type 1 investor buys whenever the type 0 investor does (in which case a buy reveals no information).
It ensures that the unique equilibrium is the fixed point with the lowest ask price: the fixed point in which \( \beta^{0,PT} = 1 \) so that no information is revealed by a CPT investor’s buy decision. The transition occurs at the lowest public belief at which the type 0 investor is willing to buy when no information is revealed by a buy decision:

\[
\left(1 - p^0\right)^{\delta - \alpha} = \lambda \left(\frac{q}{1-q}\right)^{\delta} \left(\frac{1-\mu}{2} + \mu\gamma q + \mu(1-\gamma)\right)^{\alpha}
\]  

(14)

Inspecting (12) and (14), we can see that the transition for type 0 investors always occurs at a public belief greater than one-half (the right-hand side of the equation governing \( 1 - p^0 \) is always greater than one so that we must have \( 1 - p^0 > \frac{1}{2} \)), but the transition for type 1 investors can occur at lower public beliefs.

For the decision to sell, the two equations of interest are

\[
s_t = 1 : \left(\frac{p^1}{1-p^1}\right)^{\delta - \alpha} \leq \frac{1}{\lambda} \left(\frac{q}{1-q}\right)^{\delta} \left(\frac{1-\mu}{2} + \mu\gamma(1-q) + \mu(1-\gamma)\right)^{\alpha}
\]

\[
s_t = 0 : \left(\frac{p^0}{1-p^0}\right)^{\delta - \alpha} \leq \frac{1}{\lambda} \left(\frac{q}{1-q}\right)^{\delta} \left(\frac{1-\mu}{2} + \mu\gamma q + \mu(1-\gamma)\right)^{\alpha}
\]

where I denote the probabilities with which type 0 and type 1 CPT investors sell, \( \eta^{0,PT} \) and \( \eta^{1,PT} \), respectively. Note the symmetry between the sell decision of the type 1 investor and the buy decision of the type 1 investor. The problems are in fact identical if one replaces \( p_i \) with \( 1 - p_i \). It thus follows that selling behavior is identical to buying behavior except that the transitions occur at symmetric public beliefs: \( p^0 \) for the type 1 investor and \( (1 - p^1, 1 - p^1) \) for the type 0 investor, where \( p^0 < 1 - p^1 \). At sufficiently low public beliefs, both types of investors sell. As the public belief increases, the type 1 investor first transitions to not selling (at \( p^0 \), followed by the type 0 investor (over \((1 - p^1, 1 - p^1)\)).

For \( \delta < \alpha \), the only difference is a relabeling. Type 0 investors now transition from buy to no trade in the same region that type 1 investors transition from sell to no trade in the \( \delta > \alpha \) case, and similarly for the other transitions, as illustrated in Figure 2. This duality is easily verified by comparing the inequalities that govern each transition.

For part iv), \( 1 - p^0 > p^1 \) follows from the statement in the proof of Lemma 1 that if the type 0 investor buys with positive probability, the type 1 investor must buy with probability one. Lastly, we must show \( p^0 < p^1 \). Using (12) and (14), this inequality is equivalent to

\[
\frac{1}{\lambda} \left(\frac{1-q}{q}\right)^{\delta} \left(\frac{1-\mu}{2} + \mu\gamma(1-q) + \mu(1-\gamma)\right)^{\alpha} \leq \lambda \left(\frac{1-q}{q}\right)^{\delta} \left(\frac{1-\mu}{2} + \mu\gamma q + \mu(1-\gamma)\right)^{\alpha}
\]

which is more easily satisfied for higher \( \lambda \), so take \( \lambda = 1 \). Then, the inequality becomes

\[
\frac{1-\mu}{2} + \mu\gamma(1-q) + \mu(1-\gamma) < \frac{1-\mu}{2} + \mu\gamma q + \mu(1-\gamma)
\]

\[
\iff 2\frac{1-\mu}{2} \mu\gamma(1-2q) + \mu^2\gamma(1-2q)(1-\gamma) + \mu^2\gamma^2 q(1-2q) < 0
\]

which holds for all parameterizations. □

**Proof of Lemma 2:**

Under expected utility, an investor with continuous utility function \( u(x) \) and private belief, \( b_t \), will
buy if \( b_t u(1 - A_t) + (1 - b_t) u(-A_t) \geq u(0) \)
sell if \( b_t u(B_t - 1) + (1 - b_t) u(B_t) \geq u(0) \) (15)

and otherwise abstain from trading.\(^{44}\) As in the main model, it is possible to show that because of uninformed investors, we must have \( b_t > A_t \) (favorable signal) or \( b_t < B_t \) (unfavorable signal), in which case risk-neutral investors trade according to their private information as shown in Lemma 1. Consider an investor that is not risk-neutral then, and assume she has a favorable signal (symmetric arguments hold for unfavorable signals).

(i) If risk-averse, then not trading is always preferable to selling, so herding and contrarianism are not possible: \( b_t u(B_t - 1) + (1 - b_t) u(B_t) < u(b_t(B_t - 1) + (1 - b_t)(B_t)) = u(B_t - b_t) < u(0) \). The first inequality holds because utility is strictly concave and the second because \( b_t > A_t > B_t \) with a favorable signal.

(ii) If risk-seeking, then buying is always preferable to not trading so abstention is not possible: \( b_t u(1 - A_t) + (1 - b_t) u(-A_t) > u(b_t(1 - A_t) + (1 - b_t)(-A_t)) = u(b_t - A_t) > u(0) \). The first inequality holds because utility is strictly convex and the second because \( b_t > A_t \) with a favorable signal. \( \square \)

B. Prospect Theory Parameters

When trades take place at a single price, \( p_t \), the optimal strategy of a CPT investor with a favorable private signal (from (5)) becomes

\[
\begin{align*}
\text{buy if} & \quad \left( \frac{p_t}{1-p_t} \right)^{\delta-\alpha} \geq \lambda \left( \frac{1-q}{q} \right)^{\delta} \\
\text{sell if} & \quad \left( \frac{p_t}{1-p_t} \right)^{\delta-\alpha} \leq \frac{1}{\lambda} \left( \frac{1-q}{q} \right)^{\delta}
\end{align*}
\] (16)

where \( t \) is the time of trade (after \( t - 1 \) public signal draws). For an investor with an unfavorable signal, the ratio of \( 1 - q \) to \( q \) is again inverted in each inequality. From (16), it is clear that one can define two threshold prices (that replace the transition regions that exist when there is asymmetric information) at which behavior transitions. The equilibrium is otherwise identical to that described in Theorem 1.

I now use (16) to establish that the model only has two degrees of freedom and then to find the set of pairs of threshold prices that it can support. I consider an investor with a favorable signal - the thresholds for an investor with an unfavorable signal follow by symmetry around \( p_t = \frac{1}{2} \).

The price depends only on the difference in the number of favorable and unfavorable public signals. Denoting the difference, \( k = \#\text{favorable} - \#\text{unfavorable} \), we have \( p_t = \frac{q^k}{q^k + (1-q)^k} \). (16) can then be written

\[
\begin{align*}
\text{buy if} & \quad \left( \frac{q}{1-q} \right)^{k(\delta-\alpha)+\delta} \geq \lambda \\
\text{sell if} & \quad \left( \frac{q}{1-q} \right)^{k(\delta-\alpha)+\delta} \leq \frac{1}{\lambda}
\end{align*}
\] (17)

\(^{44}\)I have normalized initial wealth to zero without loss of generality.
If we define $\lambda' \equiv \lambda^{\frac{\delta}{\delta - \alpha}}$ and $\delta' \equiv \frac{\delta - \alpha}{\delta}$, for $\delta - \alpha > 0$, we can rewrite these conditions as (for $\delta - \alpha < 0$, the inequalities are flipped):

buy if $\left(\frac{q}{1-q}\right)^{k_1 + \frac{1}{2}} \geq \lambda'$

sell if $\left(\frac{q}{1-q}\right)^{k_1 + \frac{1}{2}} \leq \frac{1}{\lambda'}$

which establishes that the model only has two degrees of freedom, $\lambda'$ and $\delta'$. One can solve for these parameters in terms of the two threshold differences in public signals, $k^0$ and $k^1$ (corresponding to $p^0$ and $p^1$, respectively), at which behavior is observed to transition, resulting in

$$\lambda' = \left(\frac{q}{1-q}\right)^{k_1 - k^0}$$  \hfill \text{(18)}

and

$$\delta' = -\frac{2}{k^0 + k^1}$$  \hfill \text{(19)}

$k_0$ and $k_1$ must satisfy several restrictions. First, herding is not possible after a single public signal. Herding after one negative public signal requires selling with a favorable signal, but the sell decision in (17) cannot be satisfied with $k = -1$ for any $\alpha > 0$. Second, from the theory, we require $p^0 < p^1$ which implies $k^0 < k^1$. Third, when $\delta > \alpha$ such that herding occurs, we must have $p^1 < 1 - p^0$, which implies $k^1 < -k^0$. However, given upper bounds on $\alpha$ and $\delta$, we must have $\delta' < 1$ which, from (19), implies the tighter restriction, $k^1 + k^0 < -2$. When $\alpha > \delta$ such that contrarianism occurs, we must have $1 - p^1 < p^0$, and therefore $-k^1 < k^0$. The restrictions on $\alpha$ and $\delta$ make $\delta'$ negative in this case so that (19) also implies $-k^1 < k^0$. I impose these restrictions when generating the set of possible trade patterns that prospect theory is capable of explaining.

Given the two transitions that best fit an individual’s behavior, we obtain $\lambda'$ and $\delta'$ from (18) and (19). To recover the primitives of the model, and even the difference, $\delta - \alpha$, requires a normalization of one of the variables. I choose to normalize $\delta = 1$ for herding types (for whom $\delta - \alpha$ is positive) and $\alpha = 1$ for contrarian types (for whom $\delta - \alpha$ is negative). These choices keep $\alpha$ and $\delta$ as close to the risk-neutral benchmark ($\alpha = \delta = 1$) as possible, while ensuring both remain weakly less than one.

C. Best symmetric model

In Section 3.2.1, I consider the ‘best symmetric’ model, a model that only imposes symmetry around a price of one-half. Here, I discuss the generality of this model in more detail.

The model imposes only the requirement that if a subject buys, sells, or abstains at a price, $p$, when she has a particular private signal, then she must sell, buy, and abstain (respectively) at a price of $1 - p$ with the opposing private signal. This requirement is very natural because of the symmetry in payoffs around a price of one-half. Buying at a price of
with a particular belief, \( b \), provides a binary gamble which returns \( 1 - p \) with probability \( b \) and \( -p \) with probability \( 1 - b \). Selling at a price of \( 1 - p \) with a belief, \( b' \), provides a binary gamble which returns \( 1 - p \) with probability \( 1 - b' \) and \( -p \) with probability \( b' \). Thus, if \( b = 1 - b' \), which is the case with Bayesian updating, the opposing actions at symmetric prices with opposing signals provide identical gambles.\(^{45}\)

Any model that respects Bayesian updating and has choices that depend only on monetary payoffs must therefore respect the symmetry requirement. In particular, all utility-based models, including those with probability weighting and reference-dependence fall into this category. Even typical models of non-Bayesian updating, such as conservativeness (overweighting the prior) or overweighting one’s private signal would respect symmetry. To capture asymmetric behavior, a model necessarily has to introduce some asymmetric primitive. The asymmetry could act through utility, such as a pure preference for buying over selling, or through beliefs, such as updating differently for favorable versus unfavorable signals. However, constructing such a model would be a post hoc exercise designed to fit the data: I’m not aware of any off-the-shelf model that microfounds such asymmetries.

D. Expectations-Based Reference Points

Here, I show that the expectations-based reference point preferences of Koszegi and Rabin (2006, 2007) are incapable of generating herding in the laboratory environment, even when extended to allow for decision weights. Koszegi and Rabin (2007) consider multiple models of reference point formation. First, they consider the case in which expectations are taken as given (they suggest that this occurs when the outcome is realized shortly after a decision is made). In this case, they define an unacclimating personal equilibrium (UPE) and an associated refinement, preferred personal equilibrium (PPE), in which the decision-maker makes a choice she is willing to follow through on. Second, they consider the case in which outcomes are realized long after a decision is made, defining a choice-acclimating personal equilibrium (CPE) where the decision-maker’s reference point has time to acclimate to the choice they made. Without taking a stance on which concept is appropriate in the context here, I show that herding is not possible in either a PPE or CPE.\(^{46}\)

A necessary condition for a decision to be either a PPE or a CPE is that it must be optimal when the outcomes it induces become the reference point. Consider the decision to buy in the context of the model. Both a PPE and a CPE require \( U(\text{buy}|\text{buy}) \geq U(\text{NT}|\text{NT}) \) and \( U(\text{buy}|\text{buy}) \geq U(\text{sell}|\text{sell}) \) where \( U(F|G) = \int \int u(w|r)dG(r)dF(w) \) is the decision-maker’s expected utility. The expectation in this expression is over both the possible wealth levels (given by the distribution \( F \)) and the possible reference points (given by the distribution \( G \)) of a reference-dependent utility over wealth, \( u(w|r) \). Koszegi and Rabin (2007) assume a reference-dependent utility function of the form, \( u(w|r) = m(w) + \mu(m(w) - m(r)) \). For

\[ b = \frac{pq}{pq + (1-p)(1-q)} = 1 - \frac{(1-p)(1-q)}{pq + (1-p)(1-q)} = 1 - b', \]

and similarly with the opposite signals.

\(^{45}\)With a favorable signal at \( p \) and unfavorable signal at \( 1-p \), \( b = \frac{pq}{pq + (1-p)(1-q)} = 1 - \frac{(1-p)(1-q)}{pq + (1-p)(1-q)} = 1 - b' \),

\(^{46}\)The UPE concept, when also allowing for probability weighting, is more difficult to work with analytically. Herding may be possible in a UPE under some conditions, but it nevertheless does not survive the PPE refinement that Koszegi and Rabin (2007) suggest.
tractability, and as assumed in many of the results of Koszegi and Rabin (2007), I assume \( m(w) = w \). Note, however, that, under Koszegi and Rabin’s assumptions on \( \mu \), the reference-dependent component of utility is S-shaped as in prospect theory. In addition, although not considered in their model, I allow for decision weights with the same functional form I assumed previously. Under these assumptions, we have

\[
U(NT|NT) = 1
\]

\[
U(buy|buy) = \frac{b_0^\delta}{(b_0^\delta + (1-b_0^\delta))^\delta} (1 - pt) - \lambda \frac{(1-b_0^\delta)^\delta}{(b_0^\delta + (1-b_0^\delta))^\delta} pt
\]

\[
+ \frac{b_0^\delta}{(b_0^\delta + (1-b_0^\delta))^\delta} (\mu(1) + \mu(-1)) \frac{\delta}{(b_0^\delta + (1-b_0^\delta))^\delta} (1 - pt)
\]

\[
U(sell|sell) = -\lambda \frac{b_0^\delta}{(b_0^\delta + (1-b_0^\delta))^\delta} (1 - pt) + \frac{(1-b_0^\delta)^\delta}{(b_0^\delta + (1-b_0^\delta))^\delta} pt
\]

\[
+ \frac{b_0^\delta}{(b_0^\delta + (1-b_0^\delta))^\delta} (\mu(1) + \mu(-1)) \frac{\delta}{(b_0^\delta + (1-b_0^\delta))^\delta} (1 - pt)
\]

In each of the latter two expressions, the first two terms correspond to the direct utility from wealth and the third term corresponds to the additional utility relative to the reference point. Importantly, in both expressions, utility relative to the reference point only derives from the cases in which the reference outcome and the actual outcome differ (e.g. for the buy case, the reference is \( 2 - pt \) and the outcome is \( 1 - pt \), or vice versa). Thus, the difference is always equal to one so that the contribution to utility from reference-dependence is the same for both a buy and a sell and drops out in the comparison of the two. The S-shaped utility function therefore plays no role, which is the reason herding is not possible.\(^{47}\)

To see this, consider an investor with an unfavorable signal. For buying to be optimal, we require

\[
U(buy|buy) \geq U(sell|sell)
\]

\[
\iff (b_0^\delta)^\delta (1 - pt) - \lambda (1 - b_0^\delta)^\delta pt \geq -\lambda (b_0^\delta)^\delta (1 - pt) + (1 - b_0^\delta)^\delta pt \]

\[
\iff (b_0^\delta)^\delta (1 - pt) (1 + \lambda) \geq (1 - b_0^\delta)^\delta pt (1 + \lambda) \]

\[
\iff \left( \frac{pt}{1-pt} \right)^{\delta-1} \left( \frac{1-q}{q} \right)^\delta \geq 1
\]

(20) requires \( pt < \frac{1}{2} \) to hold given that \( \delta < 1 \), so that an investor with an unfavorable signal can only buy at low prices, implying contrarian behavior occurs, but herding does not. Therefore, the expectations-based reference point of Koszegi and Rabin (2007) is incapable of explaining the most frequent type of behavior observed in the experiment.

\(^{47}\)Relatedly, Masatlioglu and Raymond (2016) show formally that Koszegi and Rabin’s CPE concept is equivalent to CPT without gain-loss utility (i.e. CPT with only probability weighting).
### E. Price Paths

<table>
<thead>
<tr>
<th>Market</th>
<th>Public Signal Draws</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,1,0,1</td>
<td>0.84</td>
</tr>
<tr>
<td>2</td>
<td>0,1</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>0,0,1,0</td>
<td>0.16</td>
</tr>
<tr>
<td>4</td>
<td>1,0,1,1,1</td>
<td>0.93</td>
</tr>
<tr>
<td>5</td>
<td>0,1,1</td>
<td>0.7</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td>7</td>
<td>1,1,1,1</td>
<td>0.97</td>
</tr>
<tr>
<td>8</td>
<td>0,0,0,0</td>
<td>0.03</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0.3</td>
</tr>
<tr>
<td>10</td>
<td>0,0</td>
<td>0.16</td>
</tr>
<tr>
<td>11</td>
<td>0,0,1</td>
<td>0.3</td>
</tr>
<tr>
<td>12</td>
<td>1,1,1,0</td>
<td>0.84</td>
</tr>
<tr>
<td>13</td>
<td>0,1,1,0</td>
<td>0.5</td>
</tr>
<tr>
<td>14</td>
<td>0,1,0,0</td>
<td>0.16</td>
</tr>
<tr>
<td>15</td>
<td>0,1,0</td>
<td>0.3</td>
</tr>
<tr>
<td>16</td>
<td>1,1</td>
<td>0.84</td>
</tr>
<tr>
<td>17</td>
<td>0,0,0,1</td>
<td>0.16</td>
</tr>
<tr>
<td>18</td>
<td>0,0,0,0,1</td>
<td>0.07</td>
</tr>
<tr>
<td>19</td>
<td>1,0,1,1</td>
<td>0.84</td>
</tr>
<tr>
<td>20</td>
<td>0,1,1,1</td>
<td>0.84</td>
</tr>
<tr>
<td>21</td>
<td>1,1,1</td>
<td>0.93</td>
</tr>
<tr>
<td>22</td>
<td>1,0,0,0</td>
<td>0.16</td>
</tr>
<tr>
<td>23</td>
<td>1,0,0,1</td>
<td>0.5</td>
</tr>
<tr>
<td>24</td>
<td>1,1,0</td>
<td>0.7</td>
</tr>
<tr>
<td>25</td>
<td>0,0,0</td>
<td>0.07</td>
</tr>
<tr>
<td>26</td>
<td>0,1,0,0,0</td>
<td>0.07</td>
</tr>
<tr>
<td>27</td>
<td>1,0,1</td>
<td>0.7</td>
</tr>
<tr>
<td>28</td>
<td>1,0</td>
<td>0.5</td>
</tr>
<tr>
<td>29</td>
<td>1,1,1,1,0</td>
<td>0.93</td>
</tr>
<tr>
<td>30</td>
<td>1,0,0</td>
<td>0.3</td>
</tr>
</tbody>
</table>

### F. Instructions

I have included the instructions for the MAIN treatment below. The instructions for the NO BELIEFS treatment are identical except that the second paragraph under 'A Valuable Clue' is removed.
Instructions for Trading Experiment

You are about to participate in an experiment in the economics of decision-making. In the experiment you will make decisions in several repetitions of a simulated trading game. If you follow these instructions carefully and make good decisions, you can earn a CONSIDERABLE AMOUNT OF MONEY!

In each trading game, you will be given information about an artificial stock before having the opportunity to buy or sell it (or do neither). When the game is completed, a new game will begin. There will be 30 games in all and when you are done, you will be paid according to the trading decisions you make in each game.

A screenshot of the Trading Page you will use to trade the stock is shown below. The rules of the game shown below it describe the game and the interface in detail. These rules always appear under the trading interface so that you can refer to them whenever you need to.

Please read the rules carefully and then press the Next button. You will be asked two questions to make sure you understand the rules before the games begin.
Trading Page

Game 2 of 30

Stock Price

Bin Contents
Value = 100
Value = 0

Suppose your ball is a blue ball.
Please choose an action.
- Buy
- No Trade
- Sell

Rules
Basics
The currency for the trading game is experimental currency units (ECUs) which will be converted to dollars and paid to you as a bonus at the completion of all games. The conversion rate is 100 ECUs = $0.40.

Prior to each game, the computer will randomly choose whether the stock you can trade is worth 0 or 100 ECUs, with equal probability. Because the stock's value is randomly chosen each game, there is no dependence between its value in one game and its value in any other game.
You will be given 100 ECUs at the start of each game. These ECUs are **yours to keep** if you choose not to trade. If you choose to trade, however, you will have the opportunity to earn even more money.

**The Stock's Price**

In each game, you can trade at most one unit of the stock. As shown on the Trading Page, the stock's value is represented by one of two bins of balls. If the stock is worth 100 ECUs, there are 7 blue balls and 3 green balls in the bin. If the stock is worth 0 ECUs, there are 7 green balls and 3 blue balls.

The price of the stock you can trade at is set by the computer in each game. The initial price is set to 50 ECUs reflecting the fact that there is a 50% chance the stock is worth 0 and a 50% chance it is worth 100 ECUs.

Before your turn to trade the stock, the computer will randomly draw between 1 and 5 balls from the bin that corresponds to the stock **(with replacement)**. After each ball is drawn, the computer will update the price of the stock, keeping it equal to the expected value of the stock given the information that the balls reveal (but no other information). The expected value is how much the stock is worth on average - sometimes it is worth 0 and sometimes it is worth 100 ECUs - but on average it is worth the price the computer sets given the information revealed by the balls. The stock's price can therefore provide you with information about whether the stock's value is 0 or 100 ECUs.

The price graph on the Trading Page shows the history of prices as each ball is drawn by the computer. When it is your turn, you can trade the stock at the most recent price. To see this price (or a past price) exactly, on the actual Trading Page you can hover over the point on the graph. (In the example in the screenshot, the current price is 50 ECUs.)

Given the current price, you can choose to buy, sell, or not trade the stock. If you buy the stock, you will get its value minus the price you pay for it. If you sell it, you will get its price, but will have to pay back its value. If we call the value of the stock, $V$, (either 0 or 100 ECUs), and the price of the stock, $P$, your trading profit is

- $V - P$, if you buy the stock
- $P - V$, if you sell the stock

Note importantly, that your trading profit can be positive or negative! For example, if you buy the stock at a price of 75 ECUs and it turns out to be worth 100 ECUs, you **GAIN** 25 ECUs. But, if you sell the stock for 75 ECUs and it turns out to be worth 100 ECUs, you **LOSE** 25 ECUs. Conversely, if the stock turns out to be worth 0 ECUs, you would lose 75 ECUs if you bought the stock and gain 75 ECUs if you sold it.

**A Valuable Clue**

To help you determine whether the stock is worth 0 or 100 ECUs, you will be shown a valuable clue: the color of one ball from the bin corresponding to the stock. **Only you** get to see this ball, so it does not affect the price the computer sets - you can trade at the most recent price as described above. After seeing the valuable clue, you may choose to buy, sell, or not trade the stock.
To help you interpret your valuable clue, an expert in probabilities will give you the **true** probability that the stock is worth 100 ECUs. This expert knows the price and your valuable clue, but not the stock value itself. This probability will be shown above where you are asked what trade you would like to make (not shown in the screenshot example).

**Your Trading Decisions**

In each game, you will actually make two decisions. You will be asked whether you would like to buy, sell, or not trade the stock **before** you see your valuable clue. You will first be asked how you'd like to trade if your valuable clue is a blue ball (as shown in the screenshot). After choosing buy, sell, or not trade, you will hit the Next button to confirm your trade. You will then be asked for how you'd like to trade if your valuable clue is a green ball. Then, the computer will draw a ball from the appropriate bin, show it to you, and your trading decision for that color will take place. In this way, it as if you first saw the ball and then made your trading decision.

**Your Payment**

After all 30 games have been completed, your ECUs in each game will be converted at a rate of 100 ECUs = $0.40. For example, if you have 150 ECUs after trading in a game, they are worth 150/100*$0.40=$0.60.