Investor Psychology and Asset Pricing

DAVID HIRSHLEIFER*

ABSTRACT

The basic paradigm of asset pricing is in vibrant flux. The purely rational approach is being subsumed by a broader approach based upon the psychology of investors. In this approach, security expected returns are determined by both risk and misvaluation. This survey sketches a framework for understanding decision biases, evaluates the a priori arguments and the capital market evidence bearing on the importance of investor psychology for security prices, and reviews recent models.

The best plan is...to profit by the folly of others.
— Pliny the Elder, from John Bartlett, comp.
Familiar Quotations, 9th ed. 1901.

IN THE MUDDLED DAYS BEFORE THE RISE of modern finance, some otherwise-reputable economists, such as Adam Smith, Irving Fisher, John Maynard Keynes, and Harry Markowitz, thought that individual psychology affects prices.1 What if the creators of asset-pricing theory had followed this thread? Picture a school of sociologists at the University of Chicago proposing the Deficient Markets Hypothesis: that prices inaccurately reflect all available information. A brilliant Stanford psychologist, call him Bill Blunte, invents the Deranged Anticipation and Perception Model (or DAPM), in which proxies for market misvaluation are used to predict security returns. Imagine the euphoria when researchers discovered that these mispricing proxies (such

* Hirshleifer is from the Fisher College of Business, The Ohio State University. This survey was written for presentation at the American Finance Association Annual Meetings in New Orleans, January, 2001. I especially thank the editor, George Constantinides, for valuable comments and suggestions. I also thank Franklin Allen, the discussant, Nicholas Barberis, Robert Bloomfield, Michael Brennan, Markus Brunnermeier, Joshua Coval, Kent Daniel, Ming Dong, Jack Hirshleifer, Harrison Hong, Soeren Hvidkjær, Ravi Jagannathan, Narasimhan Jegadeesh, Andrew Karolyi, Charles Lee, Seongyeon Lim, Deborah Lucas, Rajnish Mehra, Norbert Schwarz, Jayanta Sen, Tyler Shumway, René Stulz, Avanidhar Subrahmanyam, Siew Hong Teoh, Sheridan Titman, Yue Wang, Ivo Welch, and participants of the Dice Finance Seminar at Ohio State University for very helpful discussions and comments.

1 Smith analyzed how the “overweening conceit” of mankind caused labor to be underpriced in more enterprising pursuits. Young workers do not arbitrage away pay differentials because they are prone to overestimate their ability to succeed. Fisher wrote a book on money illusion; in The Theory of Interest (1930), pp. 493–494) he argued that nominal interest rates systematically fail to adjust sufficiently for inflation, and explained savings behavior in relation to self-control, foresight, and habits. Keynes (1936) famously commented on animal spirits in stock markets. Markowitz (1952) proposed that people focus on gains and losses relative to reference points, and that this helps explain the pricing of insurance and lotteries.
as book/market, earnings/price, and past returns), and mood indicators such as amount of sunlight, turned out to be strong predictors of future returns. At this point, it would seem that the deficient markets hypothesis was the best-confirmed theory in the social sciences.

To be sure, dissatisfied practitioners would have complained that it is harder to actually make money than ivory tower theorists claim. One can even imagine some academic heretics documenting rapid short-term stock market responses to news arrival in event studies, and arguing that security return predictability results from rational premia for bearing risk. Would the old guard surrender easily? Not when they could appeal to intertemporal versions of the DAPM, in which mispricing is only corrected slowly. In such a setting, short-window event studies cannot uncover the market’s inefficient response to new information. More generally, given the strong theoretical underpinnings of market inefficiency, the rebels would probably have an uphill fight.

This alternative history suggests that the traditional view that financial economists have had about the rationality of asset prices was not as inevitable as it may seem. Despite many empirical studies, scholarly viewpoints on the rationality of asset pricing have not converged. This is probably a result of strong prior beliefs on both sides. On one side, strong priors are reflected in the methodological claim that we should adhere to rational explanations unless the evidence compels rejection, and in the use of the term “risk premium” interchangeably with “mean return in excess of the risk-free rate.” For those on the opposite side, risk often comes quite late in the list of possible explanations for return predictability.

Often advocates of one approach or the other have cast the first stone out the door of their own glass house. There is, in fact, a notable parallelism among objections to the two approaches, illustrated in corresponding fashion in Table I. (Lining up each objection with its counterpart does not imply parity in the validity of the arguments.)

This survey assesses the theory and evidence regarding investor psychology as a determinant of asset prices. This issue is at the heart of a grand debate in finance spanning the last two decades. In the last few years, financial economists have grown more receptive to imperfect rational explanations. Over time I believe that the purely rational paradigm will be subsumed by a broader psychological paradigm that includes full rationality as a significant special case.

Two superb recent presentations of the asset pricing field (Campbell (2000) and Cochrane (2000)) emphasize objective external sources of risk. As Campbell puts it (p. 1516), “... asset pricing is concerned with the sources of risk and the economic forces that determine the rewards for bearing risk.” For Cochrane (p. 455), “The central task of financial economics is to figure out what are the real risks that drive asset prices and expected returns.”

In contrast, I argue here that the central task of asset pricing is to examine how expected returns are related to risk and to investor misvaluation. Campbell’s survey emphasizes the stability of the finance paradigm over the last two decades. I will argue that the basic paradigm of asset pricing is in vigorous and productive flux.
Figure 1 illustrates static asset pricing (analogous to the CAPM) when investors misvalue assets and securities. Returns are increasing with risk (measured here by CAPM beta) and with current market undervaluation of the asset. There are several potential noisy proxies for the degree of underpricing, such as price-containing variables (e.g., book-market, market value, earnings/price), measures of public mood (e.g., the weather), or actions possibly taken to exploit mispricing (e.g., recent occurrence of a stock repurchase or insider purchases). Risk and mispricing effects do not necessarily take such a simple linearly separable form (see the models described in Section IV), but it is still useful to keep the two notions conceptually distinct.

This picture is only a starting point. Just as the static risk effects of the CAPM have been generalized to intertemporal asset pricing, so the dynamic behavior of mispricing must be accounted for as well. After decades of study, the sources of risk premia in purely rational dynamic models are well understood. In contrast, dynamic psychology-based asset-pricing theory is in its infancy.

In the remainder of the introduction, I discuss market forces that can maintain or eliminate mispricing, and why we cannot dismiss mispricing on conceptual grounds. Section I of the survey presents relevant psychological biases and argues that many of the important biases grow naturally from just a few deep roots. Section II summarizes evidence on capital market and investor behavior regarding the importance of risk and misvaluation effects. Section III presents asset-pricing theories based on imperfect rationality. Section IV concludes with further directions for research.
To think about whether mispricing is viable, consider the traditional argument for rational price setting. In this account, smart traders spot dollar bills lying on the ground and grab them, which does away with mispricing. Setting aside the dynamics of wealth momentarily, the arbitrage story is incomplete in two ways. First, equilibrium prices reflect a weighted average of the beliefs of the rational and irrational traders. So long as each group has significant risk-bearing capacity, both influence prices significantly. Arbitrage is a double-edged blade: Just as rational investors arbitrage away inefficient pricing, foolish traders arbitrage away efficient pricing. Second, in some respects, all investors may be imperfectly rational. Even in the Olympics, no one runs at the speed of light; some cognitive tasks are just too hard for any of us.

The traditional argument further asserts that wealth flows from foolish to wise investors. This point carries considerable weight. Suppose that some rational individuals are immune from bias, and that all markets are liquid. Suppose that terminal dividends obey a linear factor model with $K$ systematic and $N$ idiosyncratic payoff components (I will call these systematic and idiosyncratic “factors”). An irrational investor on average trades and loses on

---

2 See, for example, Figlewski (1978), Shiller (1984), DeLong et al. (1990a), Campbell and Kyle (1993), and Shefrin and Statman (1994).
every factor that he misvalues. If the number of factors \( N + K \) is large, and
if a nontrivial fraction of them are substantially mispriced, then, on average, irrational investors lose a very large amount of money almost surely. Soon superior rationality will prevail.

Thus, as long as some investors are rational and markets are perfect, there can be substantial mispricing in only a small fraction of the \( N + K \) factors. If \( N \gg K \), then some or all of the systematic factors can be substantially mispriced, but only a small fraction of the \( N \) idiosyncratic components can be (see Daniel, Hirshleifer, and Subrahmanyam (2001a)).

On the other hand, people are likely to be more prone to bias in valuing securities for which information is sparse. This suggests that misperceptions are strongest in the dusty, idiosyncratic corners of the market place. One way to reconcile both intuitions is to recognize that there are biases that almost no one is immune to. In this case, there can be widespread idiosyncratic mispricing which only becomes apparent ex post.

Although misperceptions are probably most severe when information is sparse and arrives slowly, there is no reason to think that confusion is confined purely to idiosyncratic factors. Market timers trade based on what they perceive to be superior information about the market or about industry plays such as high-tech. Investors (whether wisely or not) purchase macroeconomic forecasts. So if investors sometimes misinterpret information, they will make systematic as well as idiosyncratic errors. Indeed, to the extent that misperceptions are conveyed through social processes, mistakes may be greatest for systematic factors along with a few well-known securities.

The fact that several empirical patterns of predictability are strongest in small (presumably less liquid) firms suggests that illiquid markets may be less efficient. This is less obvious than it sounds—the findings may result from the sparser information available about small, illiquid firms. Since arbitrage is double-edged, holding wealth constant, there is no presumption that liquidy immediately reduces mispricing. It does, however, speed the flow of wealth between smart and foolish traders, which may in the long run do so.3

It is often suggested that the expertise of hedge funds or investment banks will improve arbitrage enough to eliminate any significant mispricing. This works if foolish investors are wise enough to delegate to sound managers. However, intermediaries have incentives to serve or exploit the irrationalities of potential clients. It is not obvious that layering agency over folly improves decisions.4 So misvaluation does not require that there be frictions

---

3 Liquidity makes it easier for smart traders to arbitrage away mispricing, but also easier for foolish traders to arbitrage away efficient pricing. Barber and Odean (1999) find that traders who switch to online brokerages trade more aggressively yet subsequently perform poorly—their greater liquidity encouraged bad trades.

4 Furthermore, regardless of whether there are intermediaries, it is exactly when a security or sector becomes more mispriced that smarter investors become poorer. This weakens rational arbitrage (and strengthens irrational anti-arbitrage) in an untimely way (Shleifer and Vishny (1997) and Kyle and Xiong (2000)).
or special impediments to fundraising by smart players. Such frictions, however, can slow the flow of wealth between smart and foolish smart traders, perhaps allowing mispricing to persist longer.

When substantial mispricing is limited to a few factors and residuals, less rational investors do not necessarily lose on average to wiser ones. Investors who underestimate risk take larger long positions in risky assets, and thereby achieve higher expected returns (DeLong et al. (1990a, 1991)). It could further be argued that trading pressure by irrational investors induces cross-sectional return predictability; that these investors thereby lose money; but that on average they make up their losses by bearing more aggregate market risk. However, if overconfident investors irrationally overbuy the market, this should result in a low expected return. This does not jibe well with the equity premium puzzle of Mehra and Prescott (1985).

There are other means by which imperfectly rational individuals can earn high expected returns. Overconfident investors who buy and sell aggressively in response to valid private information signals may exploit liquidity traders more profitably than rational investors (Hirshleifer and Luo (2001)). In an imperfectly competitive securities market, overconfident traders can benefit by intimidating competing informed traders (Kyle and Wang (1997)).

Overconfident individuals are also likely to overinvest in acquiring private information, at the expense of leisure. However, what evidence we have suggests that aggressively trading individual investors do badly. Despite the ingenious explanations for profitable foolishness, it is quite plausible that, in fact, fools and their money are soon parted. Even if so, a misperception that derives from a fundamental human psychological trait can remain important for asset prices in the long term. There are two related reasons.

First, wealth is reshuffled in the process of generational succession. Second, in the process of getting rich, individuals can learn to be less rational. For example, biased self-attribution (Section I.B) causes individuals to attribute successes to their own qualities and failures to chance. As a result, losses by overconfident individuals can be offset by the rising confidence of the nouveau riche (see Daniel, Hirshleifer, and Subrahmanyam (1998) and Gervais and Odean (2001)).

It is challenging to find a source of risk to explain rationally the magnitude of cross-sectional predictability (see Section II). The challenge for the mispricing theory is to explain how irrational investors can remain important while hemorrhaging a great deal of cash. The disappearance of the size

---

6 Other means by which the imperfectly rational can do well or poorly have been described as well; see Blume and Easley (1982, 1990, 2000), Palomino (1996), Luo (1998), and Sandroni (2000).
7 See Barber and Odean (1999, 2000a, 2000b) and Odean (1999). However, most of the theoretical models imply that only investors with moderate overconfidence will do well. The data may be picking up the poor performance of the extremely overconfident.
effect in the mid-1980s and the inconsistency of the value effect in the last few years is suggestive.\(^8\)

There is a further problem. Having detected a return pattern statistically, it is hard for an investor to know whether other investors have yet detected and acted upon it. In 1984, how could an investor be sure whether other investors were overexploiting the size effect (Daniel and Titman (1999))? This uncertainty suggests that sometimes patterns of mispricing will be arbitrated away too slowly, and other times will overshoot. Conceivably the long life of the momentum effect has resulted from arbitrageurs each mistakenly fearing that others have started to trade aggressively. As Yogi Berra commented about a popular restaurant, “No one goes there any more because it’s too crowded.”

The other possible reason for persistent mispricing is that some relevant pieces of public information are ignored or misused by everyone. This can occur either because the signals are obscurely located or because our shared model of the world is just not sophisticated enough to make their relevance clear. A pricing error of this sort may disappear once a smart econometrician identifies it.

It is impossible to be comprehensive on a topic of this scope. Several important topics have been discussed in greater depth elsewhere.\(^9\) My focus is on the psychology of imperfect rationality, not psychological determinants of rational risk aversion or time preference. My benchmark for comparison throughout is the traditional asset-pricing paradigm; I do not cover market imperfections, nor models of rational bubbles.

**I. Judgment and Decision Biases**

This section describes some psychological effects that are potentially relevant for securities markets, with hints at possible explanations based upon adaptiveness.\(^{10}\) Economists have traditionally been skeptical of the varied array of seemingly arbitrary biases offered by experimental psychology. The empirical findings gain credence if we can understand what causes them. I argue here that these patterns generally derive from common roots.

\(^8\) The U.S. small firm effect was strongly positive every year during 1974 to 1983, and then was negative for six out of the next seven years; the two closing years of the millennium, which followed the publication of an important paper on “…Good News for Value Stocks” (La Porta et al. (1997)) were the worst years for value stocks since 1928, though 2000 was better.

\(^9\) I generally focus on asset market regularities involving a time horizon of at least a month, and do not consider seasonalities (for recent evidence see Hawawini, Keim, and Ziemba (2000)). Several surveys examine the equity premium puzzle in greater depth (see, e.g., Kocherlakota (1996), Campbell (1999, 2000), and Mehra and Prescott (2001)). Experiments in psychology and economics are surveyed in Camerer (1995, 1998), Bossaerts (2000), and Hertwig and Ortmann (2001).

\(^{10}\) See also the surveys of Camerer (1995), DeBondt and Thaler (1995), Rabin (1998), and Shiller (1999). There are also important literatures that build “fast and frugal” heuristics based upon ex ante considerations (Gigerenzer, Todd, and ABC-Group (1999)), and that model the decision consequences of bounds on rationality (Conlisk (1996)).
Since time and cognitive resources are limited, we cannot analyze the data the environment provides us with optimally. Instead, natural selection has designed minds that implement rules-of-thumb (“algorithms,” “heuristics,” or “mental modules”) selectively to a subset of cues (see Simon (1956)). Such heuristics are effective when applied to appropriate problems. But their inevitable biases can become flagrant when used outside their ideal domain of applicability.

Economists often argue that errors are independent across individuals, and therefore cancel out in equilibrium. However, people share similar heuristics, those that worked well in our evolutionary past. So on the whole, we should be subject to similar biases. Systematic biases (common to most people, and predictable based upon the nature of the decision problem) have been confirmed in a vast literature in experimental psychology.

There is much debate about exactly how good a job heuristics do. While psychologists such as Daniel Kahneman and Amos Tversky have made clear that heuristics can play a positive role, in the last decade, evolutionary (Darwinian) psychologists have strongly emphasized the adaptiveness of cognitive processes. In many cases biases diminish but do not vanish when probabilities are reexpressed as numerical frequencies, and when problems are posed in visual formats. However, there is no guarantee that financial decision problems will be presented to individuals in a manner that favors the most accurate decisions.

The modern environment differs greatly from the prehistoric environment of evolutionary adaption for which human cognitive mechanisms were designed by natural selection. Modern humans deal with new abstractions such as securities, money, impersonal markets, probabilities, and government and with temptations such as easy access to fats and sugars, gambling casinos, and real-time Internet trading.

The general fact that cognitive resource constraints force the use of heuristics to make decisions I will call heuristic simplification. (For cognitive resource constraints, read limited attention, processing power, and memory.) A second source of bias arises indirectly from cognitive constraints. This is that natural selection probably did not design human minds solely to make good decisions. Trivers (1985, 1991) discusses evidence that people cannot perfectly control indicators of their true internal states. This creates selection for the ability to read subtle cues such as facial expression, eye contact, posture, tone of voice, and speech tempo to infer the mental states of other individuals. In Trivers’ self-deception theory, individuals are designed to think they are better (smarter, stronger, better friends) than they really are. Truly believing this helps the individual fool others about these qualities. A third source of bias is that we are subject to emotions that can overpower reason. An evolutionary rationale for a lack of self-control is that emotions such as love and rage can act as mechanisms that allow credible commitment or

threat toward potential allies and enemies (Hirshleifer (1987); see also Frank (1988), Nesse (2001)).

I argue that heuristic simplification, self-deception, and emotional loss of control provide a unified explanation for most known judgment and decision biases. This framework can provide guidance as to which biases identified in experiments represent general mechanisms, and which are conditional side-effects.\footnote{Explanations based upon cognitive adaptiveness are subject to the objection that it is too easy to come up with "just-so" stories that fit the data ex post. However, my goal here is not to make the case that the evidence supports the adaptiveness approach (see Barkow, Cosmides, and Tooby (1992)). Rather, my point is that it is hard to make sense out of biases without a conceptual framework. Adaptiveness is about the only one we have.}

Why don’t people simply learn their way out of biased judgments? To some extent they do. One barrier is that learning is just too hard. The other barrier arises from self-deception. Individuals who think they are already competent may be slow to adjust their decision procedures (e.g., Einhorn and Hogarth (1978)).

Much of the evidence described here derives from experiments by economists and psychologists; their methods are somewhat different. Financial economists are familiar with criticisms of psychological experiments: that the stakes are low, that subjects have little experience with the experimental setting, that there is weak incentive to pay attention or tell the truth, and that publication depends on finding an effect. What may not be as familiar is that there is data addressing these issues. On the whole, training and increasing rewards and number of repetitions often reduces, but does not eliminate, biases. Lessons learned through repetition often do not carry over well across seemingly similar tasks. The well-known biases have been subjected to replication.\footnote{See, for example, Camerer (1995), Rabin (1998), and Hertwig and Ortmann (2001).} Many (though not all) of the cognitive biases are stronger for individuals with low cognitive ability or skills than for those with high ability or skills, consistent with biases being genuine errors (see Stanovich and West (2000)).

Subsections I.A and I.B consider individual biases organized by proposed causes (heuristic simplification and self-deception). Subsection I.C considers emotion and self-control, Subsection I.D discusses social interactions, and Subsection I.E discusses modeling alternatives to expected utility theory and to Bayesian updating.

A. Heuristic Simplification

A.1. Attention/Memory/Ease-of-Processing Effects

Limited attention, memory, and processing capacities force a focus on subsets of available information. Unconscious associations also create selective focus. In many studies, priming subjects with (possibly irrelevant) verbal information triggers associations that influence judgments (see, e.g., Gilovich (1981) and Higgins (1996)).
Selective triggering of associations causes salience and availability effects (e.g., Kahneman and Tversky (1973)). An information signal is salient if it has characteristics (e.g., differing from the background or from a past state) that are good at hooking our attention or at creating associations that facilitate recall. In the availability heuristic (Tversky and Kahneman (1973)), items that are easier to recall are judged to be more common. This generally makes sense, since things that are more common are noticed or reported more often, making them easier to remember. Shiller (2000b) suggested that the ease with which regular Web users can think of examples relating to the Internet revolution encouraged the market boom of the late 1990s.

One reason people are influenced by the format of decision problems is that they cannot perfectly retrieve relevant information from memory (Tversky and Kahneman (1973) and Pennington and Hastie (1988)). People underweight the probabilities of contingencies that are not explicitly available for consideration (Fischhoff, Slovic, and Lichtenstein (1978)). This suggests a kind of overconfidence (see Subsection I.B), and apparent market overreaction when unforeseen contingencies do occur.

According to self-perception theory (see Bem (1972), p. 2), “Individuals come to know their own attitudes, emotions and internal states by inferring them from observations of their own behavior and circumstances in which they occur.” The need to infer can result from memory loss or from simple lack of access to unconscious internal states. A tendency to form habits can be an optimal mechanism to address memory loss, reflecting an implicit self-perception that actions taken before probably had a good reason (Hirshleifer and Welch (2000)). Habits also economize on thinking. Habits, including the habitual adherence to self-imposed rules can also play a role in self-regulation strategies (e.g., consume only out of dividends, not principal; see Thaler and Shefrin (1981) and Shefrin and Statman (1984)).

The halo effect causes someone who likes one outstanding characteristic of an individual to extend this favorable evaluation to the individual’s other characteristics (Nisbett and Wilson (1977a)). An analogous misattribution bias could potentially cause stock market mispricing. In an efficient market, a stock being good in terms of growth prospects says nothing about its prospects for future risk-adjusted returns (which are on average zero). If people mistakenly extend their favorable evaluation of a stock’s earnings prospects to its return prospects, growth stocks will be overpriced (see Lakonishok, Shleifer, and Vishny (1994) and Shefrin and Statman (1995)).

There are ease-of-processing effects analogous to the effects of ease of retrieval. For example, the illusion of truth is the finding that people are more inclined to accept the truth of a statement that is easy to process (see, e.g., Reber and Schwarz (1999)). Familiar signal combinations (e.g., yellow with banana) are easier to perceive than unfamiliar ones (Bruner, Postman, and Rodrigues (1951)). There is a strong and robust mere exposure effect in which exposure to an unreinforced stimulus tends to make people like it more (see, e.g., Bornstein and D’Agostino (1992) and Moreland and Beach (1992)). The basis for this heuristic may be that what is familiar, being understood better, is often less risky. However, this can be taken too far, as when people
prefer to bet on a matter about which they know more than another equivalent gamble (Heath and Tversky (1991)). People also like similarity in choice of friends and mates (Berscheid and Reis (1998)). According to evolutionary psychology, people prefer familiar and similar individuals because these were indicators of genetic relatedness (e.g., Trivers (1985)). These biases suggest a tendency to prefer local investments (see also Huberman (1999)).

A literature in psychology has examined how subjects learn by observation over time to predict a variable that is stochastically related to multiple cues (see, e.g., Kruschke and Johansen (1999)). A pervasive finding is that animals and people do not achieve correct understanding of the correlation structure. Instead, cue competition occurs: salient cues weaken the effects of less salient ones, and the presence of irrelevant cues causes subjects to use relevant cues and base rates (unconditional frequencies) less. There is also learned utilization of irrelevant cues. Cue competition raises interesting questions about how information flooding through the Internet will affect misvaluation.

The learned usage of irrelevant cues comes close to magical thinking, the belief in relations between causally unrelated actions or events (as with astrology and other superstitions). A type of magical thinking called the illusion of control consists of the belief that a person can favorably influence unrelated chance events. A possible example is that people value lottery ticket numbers they select more than randomly assigned ones (Langer (1975)).

A.2. Narrow Framing/Mental Accounting/Reference Effects

Narrow framing (see Kahneman and Lovallo (1993) and Read, Loewenstein, and Rabin (1999)) involves analyzing problems in too isolated a fashion. This makes excellent sense when time and cognitive resources are limited. Many problems can be compartmentalized safely. An implication is that the form of presentation of logically identical decision problems, such as the highlighting of a different reference for comparison of outcomes can have large framing effects on choices (Tversky and Kahneman (1981, 1986)). Optimizing with respect to a problem-specific reference point, and having a direct preference over deviations (instead of over total consumption) economizes on thinking. Money illusion is another documented example of sensitivity to irrelevant description features (Shafir, Diamond, and Tversky (1997)).

By using different presentations or procedures, experimenters can elicit preference reversals. Faced with a choice between a binary lottery with a high probability but relatively low maximum payoff, versus another with lower probability and higher maximum payoff, subjects often tend to prefer the high probability lottery, yet place a higher valuation on the high-maximum-payoff lottery!14 There are also context effects in which the presence of a nonselected choice alternative affects which alternative is selected.

Mental accounting (Thaler (1985)) is a kind of narrow framing that involves keeping track of gains and losses related to decisions in separate mental accounts, and to reexamine each account only intermittently when action relevant. Mental accounting may explain the disposition effect (Shefrin and Statman (1985)), an excessive propensity to hold on to securities that have declined in value and to sell winners. Having observation of gains and losses trigger pleasant or unpleasant feelings seems a sensible mental design to motivate profitable actions. Such a mechanism may, however, be sidetracked when the individual avoids recognizing losses. Self-deception theory reinforces this argument, because a loss is an indicator of low decision ability, and a self-deceiver maintains self-esteem by avoiding recognition of such indicators.

Related arguments can explain the house money effect (Thaler and Johnson (1990))—a greater willingness to gamble with money that was recently won. The unpleasantness of a loss of recently won money may be diluted by aggregating it with the earlier gain.

Anchoring (Tversky and Kahneman (1974)) is the phenomenon that people tend to be unduly influenced in their assessment of some quantity by arbitrary quantities mentioned in the statement of the problem, even when the quantities are clearly uninformative. Some recent authors offer and test possible explanations in which the process of evaluating the anchor makes anchor-consistent arguments more accessible.¹⁵

According to expected utility theory, utility derives solely from the probability distribution of payoffs resulting from a choice. However, people seem to be regret averse in their choices (e.g., Josephs et al. (1996) and Ritov (1996)). They seem to be concerned not just that a choice may lead to low consumption, but that consumption may be lower than the outcome provided by an alternative choice.

An efficient heuristic method of comparing decision alternatives may be to line up and compare possible outcomes by state of the world (rather than evaluating the expected utility of each alternative separately and then comparing). Thus, having feelings be triggered by comparison of outcomes may be an effective mechanism for motivating good choices. Regret avoidance may also reflect a self-deception mechanism designed to protect self-esteem about decision-making ability (Josephs et al. (1996)).

Regret is stronger for decisions that involve action rather than passivity (Kahneman and Tversky (1982)), an effect sometimes called the omission bias (Ritov and Baron (1990)). Regret aversion can explain the endowment effect, a preference for people to hold on to what they have rather than exchange for a better alternative, as with the refusal of individuals to swap a lottery ticket for an equivalent one plus cash.¹⁶ The status quo bias (Samuelson and Zeckhauser (1988)) involves preferring the choice designated as the default or status quo among a list of alternatives.

Loss aversion is the phenomenon that people tend to be averse even to very small risks relative to a reference point, suggesting a kink in the utility function. This may result from the cognitive efficiency of mentally discretizing continuous variables, as reflected in the use of terms like “gain,” “break even,” and “loss,” which make the distinction between a gain and loss more salient.

A.3. The Representativeness Heuristic

The representativeness heuristic (Kahneman and Tversky (1973), Tversky and Kahneman (1974), and Grether (1980)) involves assessing the probability of a state of the world based on the degree to which the evidence is perceived as similar to or typical of the state of the world. Similarity can be viewed as an indicator of the conditional probability of the evidence given the state of the world versus other states. However, a Bayesian also takes into account heavily the prior probability of the outcomes, whereas people tend to underweight statements about unconditional population frequencies in performing conditional updating—base-rate underweighting. Furthermore, people’s perceptions of how “representative” a piece of evidence is of a state of the world may match its conditional probability poorly. For example, people tend to rely too heavily on small samples (the “law of small numbers”) and rely too little on large samples, inadequately discount for the regression phenomenon, and discount inadequately for selection bias in the generation or reporting of evidence.17 Representativeness effects have been detected in experimental markets (see Camerer (1995), Section II.C.4.)

The idea that a sample should resemble the population is often correct, especially in a large unbiased independent sample. The preceding errors amount to applying an inference too weakly within its realm of validity (large sample size) and too strongly beyond its realm of validity (small sample size). This is a natural consequence of the trade-offs involved with the design of an efficient heuristic. The resulting errors are not random: here, the error is predictable based on the sample size. The law of small numbers suggests that newly popular theories about the market drawn from recent investment experience may cause overreaction.

Misunderstanding of how randomness works can also cause a phenomenon of gambler’s fallacy. This is the belief that, in an independent sample, the recent occurrence of one outcome increases the odds that the next outcome will differ. In fact, people avoid betting on a lottery number that was a winner sometime over the preceding few days (Clotfelter and Cook (1993)).

On the other hand, use of the representativeness heuristic can cause trend chasing, because people are too ready to believe that trends have systematic causes. Statisticians refer to the clustering illusion, wherein people perceive random clusters as reflecting a causal pattern. People mistakenly believe in

17 See, for example, Tversky and Kahneman (1971); Kahneman and Tversky (1973); Nisbett and Ross (1980), Ch. 4, and references therein; Griffin and Tversky (1992); and Brenner, Koehler, and Tversky (1996b).
“hot hands” among sports players even when actual performance is very close to serially independent (Gilovich, Vallone, and Tversky (1985)). In an experimental market, consistent with gambler’s fallacy, Andreassen and Kraus (1990) find that when exogenous prices fluctuate modestly, subjects buy on dips and sell on rises. However, when a trend appears, subjects do less of this tracking and possibly switch to chasing trends. There is further evidence from experiments and from surveys that real estate and stock market investors extrapolate trends in forecasting price movements.18

A.4. Belief Updating: Combining Effects

Edwards (1968) identified the phenomenon of conservatism, that under appropriate circumstances individuals do not change their beliefs as much as would a rational Bayesian in the face of new evidence. The more useful the evidence, the greater the shortfall between actual updating and rational updating.

Having a framework for assessing biases can help when they seem to conflict. For example, conservatism implies underweighting of new evidence. Yet, if we view prior beliefs as a base rate, conservatism would seem to contradict base-rate underweighting. Perhaps conservatism is a consequence of anchoring upon an initial probability estimate. Yet, the representativeness heuristic predicts that people will extrapolate too strongly from patterns in small samples, and salience bias also causes people to overreact to certain kinds of information. Which bias do we believe?

To resolve conflicts like this requires a focus on underlying causes, and how they will operate in a particular setting. For example, self-deception can cause conservatism in a stable environment, because an individual who has explicitly adopted a belief may be reluctant to admit to himself that he made a mistake. On the other hand, if the environment is volatile, there may be no dishonor in recognizing that different beliefs are called for.

One explanation for conservatism is that processing new information and updating beliefs is costly. There is evidence that information that is presented in a cognitively costly form is weighed less: information that is abstract and statistical, such as sample size and probabilistic base-rate information. Furthermore, people may overreact to information that is easily processed, that is, scenarios and concrete examples.

The costly-processing argument can be extended to explain base-rate underweighting. If an individual underweights new information received about population frequencies (base rates), then base rate underweighting is really a form of conservatism. Indeed, base rates are underweighted less when they are presented in more salient form or in a fashion which emphasizes their causal relation to the decision problem (see Koehler (1996)). This costly-processing-of-new-information argument does not suggest that an individual will underweight his preexisting internalized prior belief. On the other hand,

if base-rate underweighting is a consequence of the use of the representa-
tiveness heuristic, there should be underweighting of priors.

Griffin and Tversky (1992) suggest that base-rate underweighting and con-
servatism, interpreted as under- versus overreaction to signals, can be un-
derstood as results of excessive reliance on the strength of information signals
and underreliance on the weight of information signals. The strength of an
information realization is how “extreme” the evidence is (in some sense) and
the weight of evidence is its reliability or precision. For example, a large
sample of conditionally i.i.d. signals has high weight. But if the preponder-
ance of favorable over unfavorable signals is modest, it has low strength.
Conservatism arises when people rely too little on high-weight evidence such
as a long sample, and base-rate neglect when people rely too heavily on
high-strength evidence such as a few signals all in one direction.

In summary, different experimental settings can lead to under- or over-
reliance on new signals; people seem to make judgments differently in dif-
ferent situations (see Grether (1992) and Payne, Bettman, and Johnson (1992)).
Given the different possible effects, invoking the name of a bias does not
provide compelling support for assuming under- or overreaction in a finan-
cial model. Further support can be provided by comparing the economic de-
cision environment of the model with the specific experimental decision setting
in which the bias was documented, and especially by running new experi-
ments that match closely the decision environment in the financial model.

Most studies of price forecasts find bias that is predictable using current
observables. For example, forecasts are often found to be adaptive, that is,
they respond partially to past forecast errors.19 Such biases are potentially
consistent either with Bayesian learning with an unknown distribution or
with overconfidence. Experimental studies involving a fixed distribution gen-
erally also yield biases, and forecasts are adaptive in most forecast experi-
ments involving endogenously determined prices as well (see Camerer (1995),
Section II.E). Consistent with overconfidence, forecasters seem to put too
little weight on the known forecasts of other forecasters.

Analyst forecasts of earnings are overoptimistic at long time horizons and
pessimistic at short horizons (e.g., Richardson, Teoh, and Wysocki (1999)).
Such biases may come from misperceptions or from agency incentives. How-
ever, we would normally expect rational agents to provide at least a positive
incremental value in their forecasting activities. There is conflicting evi-
dence as to whether stock market analysts’ forecasts of earnings do better or
worse than a time-series forecast (see the review of Kothari (2000)). A large
literature shows that real-world decision makers such as Ph.D. admission
committees or doctors do not predict outcomes as well as mechanical deci-
sion rules based on simple linear combinations of objective input measures

---

19 See the discussions in Lovell (1986) and Williams (1987), but see also Keane and Runkle
(1990). There is a similar finding for survey forecasts of macroeconomic variables (e.g., Aggar-
wal, Mohanty, and Song (1995)).
(see Camerer (1991)). This suggests that the rise of arbitrage based upon modern statistical analysis in securities markets will indeed reduce mispricing.

B. Self-Deception

The self-deception theory implies overconfidence, a very well-documented bias (as reviewed, e.g., in Odean (1998b)). An extensive literature on calibration shows that people believe their knowledge is more accurate than it really is. For example, their predictions of probabilities of events are too extreme (too high relative to the true frequency when they think the event probably will occur, too low when they think it will not). The confidence intervals they provide for quantities are too narrow, for example, 98 percent confidence intervals contain the true quantity only 60 percent of the time (Alpert and Raiffa (1982)). Experts are well-calibrated in some contexts but not others (see Camerer (1995) pp. 592–593). Experts can be more prone to overconfidence than nonexperts when predictability is low and evidence is ambiguous (Griffin and Tversky (1992)). Overconfidence is greater for challenging judgment tasks, and individuals tend to be more overconfident when feedback on their information or decisions is deferred or inconclusive.

Overconfidence is sometimes reversed for very easy items (Lichtenstein and Fischhoff (1977)). Overconfidence implies overoptimism about the individual’s ability to succeed in his endeavors. Such optimism has been found in a number of different settings (Miller and Ross (1975)). Men tend to be more overconfident than women, though the size of the difference depends on whether the task is perceived to be masculine or feminine.

Since people fail more often than they expect to, rational learning over time would tend to eliminate overconfidence. So for self-deception to succeed, nature must provide mechanisms that bias the learning process. This is consistent with self-enhancing biased self-attribution. People tend to attribute good outcomes to their own abilities and bad outcomes to external

---


21 See, for example, Lichtenstein, Fischhoff, and Phillips (1982), Yates (1990), Koren (1991), and McClelland and Bolger (1994).


23 This is not surprising on mechanical grounds; in the extreme case of perfect decision accuracy, it is possible to be under- but not overconfident about accuracy. It has been suggested that apparent overconfidence could be an artifact of the choice of questions that are not a “representative sample of the knowledge domain” (e.g., Gigerenzer, Hoffrage, and Kleinbolting (1991)), but overconfidence remains when questions are randomly selected from the knowledge domain, and has been documented in many practical choice settings (Griffin and Tversky (1992), Brenner et al. (1996a), and Soll (1996)).

circumstances. Overconfidence and biased self-attribution are static and dynamic counterparts; self-attribution causes individuals to learn to be overconfident rather than converging to an accurate self-assessment.

Self-deception also explains why there are action-induced attitude changes of the sort that motivate the theory of cognitive dissonance. In one experiment, people who chose between two products downgraded their assessments of the one they did not pick. In another, women who had to exert greater effort to gain entry to a group subsequently liked the group more. In other experiments, people who were induced with mild incentives or by request to express opinions became more sympathetic to those opinions. A tendency to be excessively attached to activities for which one has expended resources, the sunk cost effect, has been confirmed in several contexts (Arkes and Blumer (1985)). The self-deception theory suggests that a tendency to adjust attitudes to match past actions is a mechanism designed to persuade the individual that he is a skillful decision maker (see also Nel, Helmreich, and Aronson (1969) and Steele and Liu (1983)).

Similar reasoning can explain hindsight bias (e.g., Hawkins and Hastie (1990))—it helps our self-esteem to think we “knew it all along”; and the phenomenon of rationalization—constructing a plausible ex post rationale for past choices helps an individual feel better about his decision competence. People are very ready to devise and apparently believe their explanations for alleged facts about the world as well as themselves.

People tend to interpret ambiguous evidence in a fashion consistent with their own prior beliefs. They give careful scrutiny to inconsistent facts and explain them as due to luck or faulty data gathering (see Gilovich (1991) Chap. 4). This confirmatory bias can help maintain self-esteem, consistent with self-deception. Exposure to evidence should tend to cause rational Bayesians with differing beliefs to converge, whereas the attitudes of experimental subjects exposed to mixed evidence tend to become more polarized (e.g., Lord, Ross, and Lepper (1979) and Isenberg (1986)). Forsythe et al. (1992) find that individuals more subject to this confirmation bias lose money in an experimental market to those who are less subject to it. Confirmatory bias may cause some investors to stick to unsuccessful trading strategies, causing mispricing to persist.

Some general biases toward confirmation of hypotheses do not rely on self-deception. In evaluating hypotheses assigned by the experimenter about the relation of two kinds of variables (e.g., studying the night before an exam, and getting a good grade), a large literature finds that people put too much weight on confirming evidence. This involves focusing on cases in which

---

25 Langer and Roth (1975), Miller and Ross (1975), Fischhoff (1982), and Taylor and Brown (1988).
26 Festinger and Carlsmith (1959), Cooper and Fazio (1984), and Harmon-Jones and Mills (1999, Ch. 1).
27 See, for example, Nisbett and Wilson (1977b), and Ross et al. (1977b), Gazzaniga (1988) pp. 12–14.
both study and a good grade occurred, and neglecting other information (cases in which one but not the other occurred). It has been argued that such a bias is an efficient shortcut in many contexts (Klayman and Ha (1987)).

People are also biased toward seeking confirmatory information. In the famous Wason task experiments (Wason (1966)), subjects were asked to turn over cards to evaluate a hypothesis. They often turned over cards which potentially could provide instances consistent with the hypothesis, and often left unturned cards that could conclusively reject the hypothesis. A possible explanation is that positive cases are easier to process cognitively. There is evidence that people are more influenced by the information reflected in the occurrence of an event than the nonoccurrence.

C. Emotions and Self-Control

Emotions probably play a role in such traditionally rational considerations as time and risk preference, and in most or all of the effects described earlier. I discuss some further aspects of emotion here.

C.1. Distaste for Ambiguity

Choices are influenced by the structure of gambles above and beyond the overall probability distribution of consumption outcomes that the gambles provide. The Ellsberg paradoxes (Ellsberg (1961)) suggested that people are averse to ambiguity, causing irrational choices. Ambiguity aversion has been confirmed in market experimental settings. It seems to reflect a more general tendency for emotions such as fear to affect risky choices (see Peters and Slovic (1996)). As suggested by Camerer (1995), ambiguity aversion may increase risk premia unduly when new financial markets are introduced, because of the layering of uncertainty about both the structure of the economic environment and about resulting outcomes. A possible explanation for ambiguity aversion is that the obvious absence of an identifiable parameter of the decision problem may often be associated with higher risk and the possibility of hostile manipulation. This justifies a focus on missing information, but such a heuristic can go astray when there is no hostile manipulation. In a related vein, the evidence of Heath and Tversky (1991) indicates that, holding probabilities constant, people prefer gambles that give them a sense of understanding or competence.

C.2. Mood, Feelings and Decisions

Risk aversion, regret aversion, and loss aversion may reflect a calculated avoidance of unpleasant future feelings. However, mood and emotions felt today affect people’s perceptions of and choices with respect to risk.

---

28 See, for example, Jenkins and Ward (1965), Crocker (1982), and Fischhoff and Beyth-Marom (1983).
29 For example, Newman, Wolff, and Hearst (1980), and Nisbett and Ross (1980).
Mann (1992)). For example, sales of State of Ohio lottery tickets were found to increase in the days following a football victory by Ohio State University (Arkes, Herren, and Isen (1988)). More generally, people who are in good moods are more optimistic in their choices and judgments than those in bad moods (see, e.g., Wright and Bower (1992)). Bad moods are associated with more detailed and critical strategies of evaluating information (Petty, Gleicher, and Baker (1991)). The influence of mood and emotion on purchase plans and the effects of advertising have been studied by marketing researchers as well. Bodily sensations and cognitive experience (e.g., the ease and fluency of perception or recall) also affect the decision-making process (see, e.g., the survey of Schwarz and Skurnik (2001)).

Affective states (feelings or moods) contain information that individuals can use to draw inferences about the environment. However, people often attribute arousal or feelings to the wrong source, leading to incorrect judgments or misattribution biases (see, e.g., Ross (1977)). For example, people feel happier on sunny days than on rainy days, but priming them by asking them about the weather affects their judgment of how happy they are (Schwarz and Clore (1983)). Mood states tend to affect relatively abstract judgments more than specific ones about which people have concrete information. This suggests, for example, that if the weather in New York puts stock market traders in a bad mood, their pessimism may concern long-term market growth prospects rather than whether the Fed is going to lower interest rates next week.

C.3. Time Preference and Self-Control

The conventional representation of decisions over time has an additively separable utility function with exogenous, declining exponential weights. However, evidence from psychology suggests that discount rates change with circumstances. Deferring consumption involves self-control, and is therefore related to mood and feelings. There is evidence that discount rates are sometimes remarkably high, that gains are discounted more heavily than losses, that small magnitudes are discounted more heavily than large, that the framing of a choice as a delay versus an advance has a large effect on decisions, that time preference differs greatly in different decision domains (e.g., money versus health), and that visceral influences such as pain or hunger affect intertemporal choices.

The exponential specification is time consistent. However, experimental studies suggest that people and nonhuman animals are time-inconsistent. Specifically, they tend to discount a deferral of consumption from date $t$ to

---

31 See for example, Wilson and Schooler (1991), and Clore, Schwarz, and Conway (1994).
32 Clore et al. (1994) and Forgas (1995).
33 See, for example, the discussions of Loewenstein and Prelec (1992), Loewenstein (1996, 2000), and Chapman (1998).
$t + 1$ more heavily as date $t$ approaches, consistent, for example, with a hyperbolic form for discount rates.$^{34}$ This causes choice reversals even when no new information arrives. Hyperbolic discounting has been disputed.$^{35}$ Nevertheless, recent economic studies have applied time-inconsistent discounting to a wide range of issues including savings, liquidity premia, and the equity premium puzzle.$^{36}$

D. Social Interactions

Financial economists have borrowed more from the psychology of the individual than from social psychology. Financial theorists have examined how information is transmitted by prices, volume, or corporate actions. However, person-to-person and media contagion of ideas and behavior also seems important. People tend to conform with the judgments and behaviors of others, as documented in the famous length-estimation experiments of Asch (1956). A meta-analysis of 133 related studies (Bond and Smith (1996)) confirmed the conformity effect, which is, however, history and culture dependent. There are rational informational reasons to learn by observing the actions of others.$^{37}$ However, a fully descriptive analysis will have to encompass imperfect rationality (see, e.g., Ellison and Fudenberg (1995)).

Conversation is critical in the contagion of popular ideas about financial markets, as emphasized by Shiller (2000a).$^{38}$ In a survey of individual investors, Shiller and Pound (1989) found that almost all of the investors who recently purchased a stock had their attention drawn to it through direct interpersonal communication. The influence of conversation on trading may arise from individuals’ overconfidence about their ability to distinguish pertinent information from noise or propaganda; examples of large price movements triggered by Internet chat comes to mind.

As discussed by Shiller (1999), owing to limited attention, people tend to pay much more attention to ideas or facts that are reinforced by conversation, ritual, and symbols. In consequence, culture becomes an important determinant of behavior, and expression of ideas can be self-reinforcing. Kuran and Sunstein (1999) describe the process of belief formation as leading to “availability cascades,” wherein an expressed perception is perceived to be more plausible as a consequence of its increased availability in public discourse.

$^{37}$ See, for example, Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992); on the possibility of informational cascades in securities markets, see Avery and Zemsky (1998) and Lee (1998).
$^{38}$ See the conversational learning models of Banerjee and Fudenberg (1999) and Cao and Hirshleifer (2000).
Conversation pools information surprisingly poorly. Groups of people tend to talk much more about information signals that they already share than individual-specific signals (Stasser, Taylor, and Hanna (1989)). As a result, groups sometimes fail to detect patterns that are discernable by combining individual-specific signals (Stasser and Titus (1985)). Environmental pressures such as crowding and unusual circumstances cause group members to experience “cognitive overload,” and rigid thinking (adherence to habitual behaviors; see Argote, Turner, and Fichman (1988)).

When communicating information, people tend to sharpen and level, that is, emphasize what they construe to be the main point, and deemphasize qualifying details that might confuse this point. This is necessary for clarity given cognitive constraints (Anderson (1932) and Allport and Postman (1947)), but tends to cause listener beliefs to move to extremes. A closely related point is that causes tend to be oversimplified, distorting listener beliefs. There are also systematic message distortions related to a desire to be entertaining or to manipulate the listener (see Gilovich (1991), Chap. 6). These facts point to the need for analysis of conversation and rumors in securities markets.

The fundamental attribution error (Ross (1977)) is the tendency for individuals to underestimate the importance of external circumstances and overestimate the importance of disposition in determining the behavior of others. In a financial context, such a bias might cause observers of a repurchase to conclude that the CEO dislikes holding excess cash rather than that the CEO is responding to market undervaluation of the stock. This would suggest market underreaction to corporate events.

People mistakenly believe that others share their beliefs more than they really do, the false consensus effect (e.g., Ross, Green, and House (1977a); Welch (2000) reports a false consensus effect among financial economists about the equity premium). Self-deception may encourage this by making the individual reluctant to consider the possibility that he is making a deviant error. False consensus may also result from availability (since like-minded people tend to associate together). The curse of knowledge (Camerer, Loewenstein, and Weber (1989)) is a tendency to think that others who are less informed are more similar in their beliefs to the observer than they really are.

E. Modeling Alternatives to Expected Utility and to Bayesian Updating

Expected utility theory has dominated financial modeling because it captures rational decision making elegantly. However, the paradoxes of Allais (1953) and subsequent confirmations showed systematic violations of expected utility; people seem to be influenced by “irrelevant alternatives.” Further violations have multiplied. Evidence of systematic preference reversals suggests that choice may not be well described by maximization of a utility function at all. A less radical departure from the traditional approaches is to consider alternative objectives (Camerer (1995, 1998) provides an in-depth treatment). Camerer discusses generalizations that involve functional forms
on probability weightings and utility functions, in some cases explicitly derived from modified axioms of choice.

In prospect theory (Kahneman and Tversky (1979) and Tversky and Kahneman (1992)), individuals maximize a weighted sum of “values” (analogous to utilities), where the weights are functions of probabilities (instead of true probabilities). Extremely low probabilities are treated as impossibilities, and extremely high probabilities as certainties. In contrast, very (but not extremely) low probabilities are overestimated, and very (but not extremely) high probabilities are overestimated. For intermediate probabilities, the weighting function increases with a slope less than one. The value function is kinked at the “reference point” (loss aversion). The value function is concave to the right of the reference point and convex to the left, reflecting risk aversion among gambles that involve only gains and risk seeking among gambles involving only losses.

The advantage of this approach is that it can capture many of the known patterns of individual choice under risk, as well as financial regularities (see, e.g., Camerer (1998) and Shiller (1999)). Indeed, Camerer (1998) argues that a form of prospect theory fits the data better than either expected utility theory or the other generalizations that have been proposed.

Several other generalizations of time-additively separable expected utility have been applied to asset pricing issues, especially the equity premium puzzle. Epstein and Zin (1989) developed a class of intertemporal utility functions that allow for nonadditivity and nonexpected utility behavior. Priming is a phenomenon in which exposure to a stimulus affects a subject’s later response to further presentation of the same or a related stimulus. Evidence of priming effects does not tell us how people react to repeated consumption choices (self-administered stimuli of a sort), but is broadly suggestive that past consumption levels may influence how people respond to future consumption levels. Such dependence is reflected in habit formation preferences (Sundaresan (1989) and Constantinides (1990)), in which the utility derived from current consumption also depends on a habitual level of consumption.

Gilboa and Schmeidler (1995) offer a case-based decision theory which, unlike expected utility theory, is not based on evaluating outcomes and their probabilities. A case is a menu of decision options. Choices are evaluated based on outcomes of past choices and how similar those choices are to those in the current menu.

The evidence on heuristics and biases also suggests that Bayesian updating is not fully descriptive of human behavior. However, Bayes theorem is nonarbitrary, which is a useful discipline for modeling. Some recent models describe updating based on self-attribution bias and confirmatory bias.40

39 First-order risk averse preferences (Epstein and Zin (1990)), like loss aversion, involve a utility function that depends on a reference point, and in which there is nontrivial aversion even to small risks. In the case of disappointment aversion (Gul (1991)), investors weigh outcomes that are worse than the certainty-equivalent outcome more heavily than favorable outcomes.

II. Evidence of Risk and Mispricing Effects

I classify the evidence bearing on asset mispricing into five categories: (1) return predictability; (2) the equity premium puzzle; (3) evidence as to whether firms take actions in response to mispricing; (4) whether firms take actions in order to create mispricing; and (5) evidence of investment errors. My emphasis here is on findings that have received confirmation over time and location. However, such consistency is not a prerequisite for a pattern to be interesting. If widespread and fairly stable patterns of mispricing exist, then almost surely transient and situation-specific ones do too.

A. Predictability of Security Returns

Return predictability research is haunted by the specter of data mining. Some of the patterns described here are probably just vagaries of chance. However, predictability is a generic prediction of modern asset-pricing theories. So cautious skepticism rather than profound suspicion is called for.

Most of the patterns of return predictability summarized here have dual (and duelling) explanations based on either risk premia or mispricing. Empirical papers on predictability often interpret psychological explanations naively. Several authors interpret evidence that factor loadings or aggregate conditioning variables can capture predictability as counter to the psychological approach. But the psychological approach recognizes that investors should care about factor risk. To attribute a return pattern to rational factor pricing requires not just a finding that factors matter, but measurement of whether expected returns are commensurate with the relevant risks. Furthermore, the psychological approach predicts that factors, not just residuals, will be mispriced. The conditioning variables and the variables used to identify factors, such as aggregate dividend yield, the term premium, the default premium, book/market, and size, are very natural proxies for factor misvaluation, as will be discussed.

A.1. Predictability Based Upon Factor Risk Measures

I focus here on CAPM beta and the factor loadings of Fama and French (1993). A positive univariate relation of beta with expected returns is found in most studies, but depends on the country, time period, empirical implementation, and form of the CAPM being tested. Beta has incremental

\footnote{This topic is vast; for recent reviews of different aspects of the evidence pertaining to mispricing, see Fama (1991, 1998), Kothari (2000), Daniel, Hirshleifer and Teoh (2001), and Lee (2001). I do not discuss actions by outsiders such as mutual funds to exploit predictability.}

power to predict future returns after controlling for market value and/or fundamental/price ratios in some studies but not others.\footnote{See, for example, Fama and French (1992, 1996a), Handa et al. (1993), Kothari et al. (1995), Jagannathan and Wang (1996), Kim (1997), Knez and Ready (1997), Heston et al. (1999), and Kothari and Shanken (2000).

This insight has been applied to stocks by Ball (1978), Keim (1988), and Berk (1995).


A.2. Predictability Based Upon Price and Benchmark Value Measures

A natural way to identify mispricing is to compare an asset’s price to a related value measure. A remarkably consistent empirical pattern is that almost any such pairing that researchers try predicts future returns in the right direction—the “cheap” security on average appreciates relative to a risk-adjusted benchmark, or relative to an “expensive” security. Efficient market fans will conclude, however, that the security is cheap because it is riskier, and that the risk adjustment is misspecified.\footnote{See Swaminathan (1996).}

In several cases, the market value of a parent firm has been substantially less than one of its parts, and managers undertook transactions apparently suitable for exploiting the overpricing of a division.\footnote{See Fama and French (1992), Jagannathan and Wang (1996), Kim (1997), Knez and Ready (1997), Heston et al. (1999), and Kothari and Shanken (2000).}

Closed end funds often trade at discounts and premia relative to net asset value; these discounts are sometimes traded at different prices by different clienteles (Rosenthal and Young (1990) and Froot and Dabora (1999)).

A short-term yield provides a value benchmark for a long-term bond. Discrepancies between long- and short-term yields positively predict the holding period returns on long-term bonds.\footnote{See, e.g., Engel (1996).}

Bonds denominated in different currencies provide mutual benchmarks. Investing in a country's bonds that have recently become cheaper (higher nominal yield) relative to another country's bonds on average earns higher returns—the forward premium puzzle (see, e.g., Engel (1996)).


For the stock market as a whole, high fundamental/price ratios (dividend yield or book/market) predict future in-
dex returns in the United States and internationally in several, though not all, studies.\textsuperscript{49} A better predictor of cross-sectional and aggregate returns can be formed by normalizing price with earnings-based indices of fundamental value.\textsuperscript{50} Market returns are also predictable based on term and default spreads.\textsuperscript{51}

Size and value portfolios are associated with a factor or factors distinct from the stock market portfolio.\textsuperscript{52} The loadings on three factors based on size, value, and the market predict the returns on portfolios sorted on various characteristics, but do not explain short-term momentum; a global two-factor model predicts international returns (Fama and French (1996b, 1998)).

Several studies report very high Sharpe ratios achievable based on cross-sectional value effects,\textsuperscript{53} a point reinforced by low international correlations of some size and value strategies (Hawawini and Keim (1995)). This raises the question of whether the implied variability of marginal utility across states under rational asset pricing is implausibly high (see Hansen and Jagannathan (1991)). Chen (2000) finds that book-market and momentum-based portfolios do not contain enough information about future returns on aggregate wealth to be strongly priced as state variables in a Merton ICAPM.

Fama and French (1995) suggest that size and book-market factors may be correlated with harms suffered by individuals when firms are distressed. Differing conclusions have been drawn about the association of size and book-market with distress.\textsuperscript{54} The book-market effect remains strong after controlling for distress (Griffin and Lemon (2001)). The voluntary allocation by employees of personal retirement funds into shares of their own firms (Be- nartzi (1997)) opposes the distress–risk hypothesis.

Conclusions differ as to whether “characteristics” (size, book-market) or factor loadings do a better job predicting returns.\textsuperscript{55} Perhaps the most compelling evidence for expectational errors is that, after portfolios are formed, growth stocks on average respond very negatively to subsequent earnings announcements for several years, and value stocks do not (La Porta et al. (1997) and Skinner and Sloan (2000)).


\textsuperscript{51} See Keim and Stambaugh (1986), Campbell (1987), and Fama and French (1989).

\textsuperscript{52} See Fama and French (1993, 1995), and Liew and Vassalou (2000). This of course does not guarantee that the loadings on these factors are priced separately from market beta; for example, under the CAPM they would not be.


\textsuperscript{54} Chan and Chen (1991), Shumway (1996), Chen and Zhang (1998), and Dichev (1998).

\textsuperscript{55} See Daniel and Titman (1997), Jagannathan et al. (1998), Lewellen (1999), Davis et al. (2000), and Daniel et al. (2001).
A.3. Predictability Based Upon Past Returns: Momentum and Reversal

In many asset and security classes internationally, there is positive short-lag autocorrelation and negative long-lag autocorrelation.56 Cross-sectionally, U.S., European, and emerging market stocks that have done very well in the recent past (about 3 to 12 months) tend to do well over the next month.57 Long term reversals in the cross section are documented by DeBondt and Thaler (1985).58 Momentum is stronger in small firms, growth firms, firms with low analyst following, and in the security-specific (nonmarket) component of stock returns.59 Volume interacts with momentum in predicting future returns, suggesting a possible cycle of overreaction and correction (Lee and Swaminathan 2000b). Chordia and Shivakumar (2000) report that momentum profits can be captured based on security sensitivities to a few aggregate variables (see also Ahn, Conrad, and Dittmar 2000). Lewellen (2000) provides evidence of negative autocorrelation and cross-serial correlation in industry and size portfolios, consistent with negative market autocorrelation during the study’s time period.60

Past winners earn substantially higher returns than do past losers at the dates of quarterly earnings announcements occurring in the seven months following portfolio formation.61 This is surprising from a rational risk perspective, because high momentum firms should become less leveraged and less risky. Also, firms with extremely low returns over the several months are having trouble, so the distress factor view of value effects suggests that negative momentum firms should earn high future returns.

A.4. Predictability Based Upon Public Versus Private News Events

Several event studies have documented abnormal returns subsequent to the event date. One explanation, event selection, is that a firm’s decision
whether and when to engage in the event depends on whether there is market misvaluation. A second possibility, manipulation, is that around the time of the action, the firm reconfigures other information reported to investors in order to induce misvaluation.

There is evidence suggesting that both selection and manipulation occur. Regarding selection, a remarkable pattern emerges from studies of discretionary corporate events (actions chosen by management or other potentially informed parties). The average abnormal stock return in the three to five years subsequent to the event has the same sign as the event-date stock price reaction. I call this regularity postevent return continuation.\footnote{Events for which this has been found include stock splits (Grinblatt, Masulis, and Titman 1984), Ikenberry, Rankine, and Stice (1996), Desai and Jain (1997b), and Ikenberry and Ramnath (2000); tender offer and open market repurchases (Lakonishok and Vermaelen 1990 and Ikenberry, Lakonishok, and Vermaelen (1995, 2000); equity carveouts (Vijh 1999)); spinoffs (Cusatis, Miles, and Woolridge (1993) and Desai and Jain (1997a)); accounting write-offs (Bartov, Lindahl, and Hicks (1998)); analyst earnings forecast revisions (Chan et al. (1996) and Lin (2000a, 2000b)); analyst stock recommendations (Groth et al. (1979), Bjerring, Lakonishok, and Vermaelen (1983), Elton, Gruber, and Gultekin (1984), Womack (1996), Michaely and Womack (1999), Krische and Lee (2000), and Barber et al. (2001)); dividend initiations (Michaely, Thaler, and Womack (1995) and Boehme and Sorescu (2000)); dividend omissions (Michaely et al. (1995)); seasoned issues of debt (Spiess and Affleck-Graves (1999)); seasoned issues of common stock (Loughran and Ritter (1995), Spiess and Affleck-Graves (1995), Cornett et al. (1998), Teoh et al. (1998b), Jegadeesh (1999), Foerster and Karolyi (2000), but see Kang, Kim, and Stulz (1999)); public announcement of previous insider trades (Rozeff and Zaman (1988) and Seyhun (1988)); and venture capital share distributions (Gompers and Lerner (1998)). The hypothesis has not been tested for IPOs since we do not observe the price reaction to the announcement that an IPO will occur. The pattern does not hold for exchange listing (McConnell and Sanger (1987), Dharan and Ikenberry (1995), and Foerster and Karolyi (1999)) and private placements (Hertzfeld et al. (1999)), which may involve informed discretion on the part of the buying as well as the selling party.}

The evidence that has appeared since this postevent return continuation hypothesis was proposed by Daniel et al. (1998) has generally been supportive of new time periods and events. There has been little study of postevent performance for events that are not taken at the discretion of management or analysts with incentives to react to mispricing. However, Cornett et al. (1998) find that there is postevent continuation when bank stocks issue equity, except when equity issuance is forced by reserve requirements.

Fama (1998) argues that these return patterns are sensitive to empirical methodology. Several recent studies have concluded that there is limited or no underperformance of new issue firms.\footnote{See Brav, Geczy, and Gompers (2000), Eckbo and Norli (2000), Eckbo, Masulis, and Norli (2000), Gompers and Lerner (2000), and Mitchell and Stafford (2000).} However, some recent methods minimize the power to detect misvaluation effects (Loughran and Ritter (2000)). Jegadeesh (1999) reports large post-SEO underperformance even relative to several (excessively) stringent return benchmarks.

The argument that post-IPO underperformance is eliminated by an appropriate benchmark is counterintuitive, because it amounts to saying that IPO firms have unusually low risk. Risk reduction may justifi...
turn benchmark for SEO firms, but risk increases would seem to imply a higher benchmark after debt issues or bond rating downgrades, making the underperformance after these events64 even stronger. Poor post downgrade performance also opposes the distress-risk-factor theory of return predictability. New issue firms perform especially badly at subsequent earnings announcement dates, which is hard to interpret as a negative risk premium.65

Irrelevant, redundant, or old news affects security prices when presented saliently.66 These demonstrable examples of mispricing suggest that less blatant mispricing may occur routinely. Little of stock price or orange juice futures price variability has been explained empirically by relevant public news.67 Historical crashes and speculative episodes are often hard to explain in terms of fundamental news.68 Allen (2001) provides examples suggesting that bubbles have major economic consequences, and argues that agency problems among financial institutions may cause bubbles.

Several studies explore fundamental trends and subsequent returns. Cash or earnings surprises are followed by positive abnormal returns in the short run, and perhaps negative abnormal returns in the long run.69 Investors also seem to extrapolate fundamentals in options and in football betting markets (Avery and Chevalier (1999) and Poteshman (2001)).

A.5. Predictability Based Upon Mood Proxies

Environmental factors that influence mood are correlated with stock price movements. Kamstra, Kramer, and Levy (2000a) find that a deterministic variable, changes to and from daylight saving time, disrupts sleep patterns, and is related to stock returns.70 A stochastic variable, cloud cover in the city of a country’s major stock exchange, is associated with low daily stock index returns in a joint test of 26 national exchanges as well as in the United States (Hirshleifer and Shumway (2001) and Saunders (1993)).

65 Jegadeesh (1999) and Denis and Sarin (2000); see also Ikenberry and Ramnath (2000).
68 See, for example, Cutler et al. (1989), Seyhun (1990), Shiller (2000b) Ch. 4; for a mainly rational perspective on the Dutch tulip bulb boom, see Garber (1989).
69 On short run postearnings announcement drift, see, for example, Bernard and Thomas (1989, 1990). For poor long-lag performance, see DeBondt and Thaler (1987), Lakonishok et al. (1994), Lee and Swaminathan (2000a), but see also DeChow and Sloan (1997), and Daniel and Titman (2000).
70 Kamstra, Kramer, and Levi (2000b) examine the relation of another deterministic variable, seasonal shifts in length of day, to returns in several countries.
B. Equity Premium and Risk-free Rate Puzzles

The equity premium puzzle is that U.S. equity market returns are high relative to risk, implying high levels of risk aversion and so a low elasticity of intertemporal substitution in consumption. This in turn implies very high real interest rates to induce individuals to accept lower consumption now than in the future (consistent with historical growth in consumption; see Weil (1989)).

C. Actions Possibly Taken in Response to Mispricing

Corporations buy and sell shares in a way that is correlated with possible measures of market mispricing. The amount of financing and repurchase varies widely over time in an industry-specific way. Merger bids, which often rely on equity financing, are also prone to booms and quiet periods by industry. New closed-end funds are started in those years when seasoned funds trade at small discounts or at premia relative to net asset value (Lee, Shleifer, and Thaler (1991)), and tend to be issued at a premium (plus commission) before reverting to a discount in the aftermarket (Peavy (1990)).

D. Actions Possibly Taken to Create Misvaluation

Firms sometimes make accounting adjustments (accruals) to boost their earnings relative to actual cash flow. These adjustments are publicly disclosed in firms’ financial statements. When accruals are abnormally high, stocks on average subsequently experience poor return performance. Managers boost accruals at the time of new IPO and seasoned equity issues (Teoh et al. (1998a, 1998b)). Greater earnings management in IPOs and in SEOs are associated with more optimistic errors in analyst earnings forecasts, and with more adverse subsequent long-run abnormal stock returns.

Managers adjust earnings to meet threshold levels such as zero, past levels, and levels forecast by analysts (DeGeorge, Patel, and Zeckhauser (1999)). Possibly under the influence of management, stock analysts on average “walk down” their forecasts from overly optimistic levels at long horizons to pessimistic forecasts that firms are likely to beat by year-end (Richardson et al. (1999)).

References:
71 See Shiller (1982), Hansen and Singleton (1983), Mehra and Prescott (1985), and Hansen and Jagannathan (1991). Purely rational explanations have been offered based upon learning (Brennan and Xia (2001)), luck (Fama and French (2000)), selection bias in the focus of academic attention (Brown, Goetzmann, and Ross (1995)), borrowing constraints (e.g., Constantinides, Donaldson, and Mehra (2000)), and non-stock-market income shocks (e.g., Constantinides and Duffie (1996) and Heaton and Lucas (1996)).
72 See, for example, Jindra (2000), D’Mello and Shrof (2000), and Dittmar (2000); Korajczyk, Lucas, and McDonald (1991) provide a possible rational explanation for this phenomenon.
74 See Teoh et al. (1998a, 1998b) and Teoh and Wong (2001).
E. Quality of Information Aggregation

In contrast with early classic work on experimental markets, the thrust of much experimental market research in the late 1980s and 1990s is that in only a slightly more complicated environment, information is not aggregated efficiently (see, e.g., the survey of Libby, Bloomfield, and Nelson (2001) and Sunder (1995)). Presumably this is because confounding effects make it harder for investors to disentangle the reasons behind the trades of others (see, e.g., Bloomfield (1996)).

F. Investor Behavior

Portfolio theory suggests that (apart from transaction costs) everyone should participate in all security markets. But even now, many investors neglect major asset classes. Nonparticipation may derive from salience bias, or from mere exposure (familiarity) effects. Investors are subject to a strong bias toward investing in stocks based in their home country and in their local region. Employees invest heavily in their own firm's stock and perceive it to have low risk (Huberman (1999)). The degree to which they invest in their employer's stock does not predict the stock's future returns (Benartzi (1997)). There is also experimental evidence that investors sometimes fail to form efficient portfolios and violate two-fund separation.

Several, though not all, studies of investor behavior in natural and experimental markets report evidence consistent with a disposition effect—a greater readiness to realize gains than losses. Certain groups of investors change their behaviors in parallel (“herding”), in some cases engaging in momentum (or positive feedback) trading and in other cases in contrarian trading. Similar behavior is not irrational per se, but some groups of investors do poorly.

People (especially males) seem to trade too aggressively, incurring higher transactions costs without higher returns. Furthermore, traders in experimental markets do not place enough weight on the information and actions of others (Bloomfield, Libby, and Nelson (1999)). Both findings are consistent with overconfidence. In experimental markets, as in psychological ex-

---

77 See, for example, Shefrin and Statman (1985), Ferris, Haugen, and Makhija (1988), Odean (1998a), Weber and Camerer (2000), and Grinblatt and Keloharju (2001b); but see also Rangelova (2000).
periments, investors and prices are more prone to overreacting to unreliable than to reliable information.  

Investors not infrequently make flagrant errors, such as failing to exercise in-the-money options at expiration, and apparently failing to exploit arbitrage opportunities (Longstaff, Santa-Clara, and Schwartz (1999) and Rietz (1998)). In retirement fund contribution decisions, there is evidence that people are strongly subject to status quo bias, diversify naively by dividing their contributions evenly among the options offered, and appear to naively extrapolate past return performance.

III. Asset-pricing Theories Based on Investor Psychology

A. The Origins of Psychology-based Asset Pricing

The evidence in the preceding section presents challenging puzzles to be explained. It is natural to consider whether failures of rationality explain these patterns. As mentioned in footnote 1, for many years, economists theorized without apology about the effects of imperfect rationality. Upon its ascendancy, there were critics of the fully rational paradigm. The most fully developed critiques and alternative approaches arose from the work of Herbert Simon and colleagues at Carnegie Mellon University, and from experimental psychologists such as Daniel Kahneman and Amos Tversky. Simon offered a theory of bounded rationality, in which limited calculating power and the complexity of decision problems prevent fully rational decisions. Simon (1955) proposed that instead individuals engage in “satisficing,” the choice of solutions that are good enough. These ideas were further developed to explore organizational slack, the supoptimal use of resources by firms (see Cyert and March 1963).

Bounded rationality implies a need for simple heuristics for making decisions, suggesting that computational simulation of heuristic agents may be a fruitful line of inquiry for finance. The costliness of evaluating contingencies suggests that the market may undervalue complex securities, and that imperfect rationality affects the introduction and success of new securities. Simon also studied the optimal design of cognitive mechanisms given resource constraints. This design stance toward the human mind provides a way of thinking about the source of biases identified in psychological experimental studies.

The experimental psychology literature (see Section I) attempts to identify and describe heuristics and biases. Researchers from this intellectual lineage recognize that heuristics are designed to be adaptive, but the focus is on identifying the conditions that can cause heuristics to work poorly. The experimental psychology evidence has been applied more directly in finance

80 See Bloomfield et al. (2000, 2001).
81 See Benartzi (1997), Madrian and Shea (2000), and Benartzi and Thaler (2001).
than have general notions of bounded rationality, because a priori reasoning about the effects of cognitive design constraints remains speculative until empirically confirmed. However, as argued in Section I, the design stance remains important because it helps researchers evaluate alternative heuristics that are potentially consistent with the experimental findings.

The application of psychological evidence to problems in economics was led by Richard Thaler and colleagues starting in the 1980s. Psychological ideas relating to overreaction stimulated the collection of empirical evidence on stock price serial correlation and volatility (see Section II). The 1986 publication of Fischer Black’s (1986) AFA presidential address on “Noise,” and the 1987 stock market crash further increased academic interest in imperfectly rational trading.

Some pioneering analytic models captured imperfect rationality in asset markets by including mechanistic traders who either make pure noise trades, or positive feedback trades in which new purchases are an increasing function of past price moves.82 This was an efficient way to illustrate some crucial insights about survival, arbitrage, and pricing. However, in full generality, the mechanistic modeling approach is very elastic. If noise trades can be arbitrarily correlated with other economic variables, any return pattern can be explained. The economic content of mechanistic trader models comes from the choice of assumptions on trades to reflect facts about psychology or trading. In the hope of being more accurately predictive, recent research has explicitly modeled how decision making occurs in a way that reflects psychological biases.

In a specific investment setting, it can be hard to judge which documented psychological bias is relevant. This creates an extra degree of freedom for model-mining not present in the purely rational approach. Thus, even more than for purely rational theories, a psychological theory becomes more persuasive if it explains a range of empirical patterns in different contexts, and generates new implications.

The next subsection starts with models of the simple statics of mispricing and correction. Models of dynamics follow in Subsection III.C. A static setting can address how risk and mispricing determine the cross section of expected returns. Mispricing proxies capture long-term misvaluation and correction. Models of dynamics can describe intertemporal patterns, such as a shift from underreacting to overreacting to a stream of news, or a pattern of overreacting and then overreacting even more. Thus, dynamic analyses can address patterns of short- versus long-term return autocorrelations. Subsection III.D discusses how to empirically distinguish psychology-based pricing theories.

B. Static Asset Pricing

I consider static models based upon either limited attention/participation, or overconfidence. Merton (1987) analyzed the cross section of security re-

82 See Frankel and Froot (1986, 1990), Cutler et al. (1990), and DeLong et al. (1990a, 1990b).
returns in a static asset-pricing model with exogenous nonparticipation. Such nonparticipation can be viewed as reflecting limited attention, preference for the familiar, and salience effects. The key implication of the model is that neglected stocks earn abnormally high expected returns.

Some recent static analyses of psychology and security returns are based on investor overconfidence. Financial analysts and investors differ in their skill at acquiring information through means such as interviewing management, analyzing financial statements, and Internet chat. An investor who overestimates his ability to do so will underestimate his errors in forecasting value. Thus, as in Kyle and Wang (1997), in these models an overconfident investor overestimates the precision of his information signals.

Odean (1998b) studies the statics of overconfidence when there is a single risky security. When price-taking investors think the signal is more accurate than it really is, the market price overreacts to the signal. Eventually, when the true state of the world resolves, the price corrects. This pattern of overreaction and reversal causes excess price volatility, and negative long-run return autocorrelation.

Instead of a general tendency to overestimate signal precision, in Daniel et al. (1998), investors are only overconfident about private information signals. This reflects the notion that an investor’s self-esteem is tied to his own ability to acquire useful information. Individuals receive a private signal, and subsequently update based on an inconclusive public signal. In the static version of the model, investor confidence is fixed. Managers may selectively undertake good news activities such as a stock split or repurchase at least partly in response to market undervaluation of the firm, and other activities such as new issue when the firm is overvalued.

Since investors overreact to private signals, returns on private information arrival dates tend to reverse. In contrast, for selective public events, the model implies postevent continuation of stock returns: selective events associated with positive (negative) average event-date reactions are also associated with positive (negative) average postevent long-run abnormal returns. Intuitively, when the firm (or another party) takes a public action in opposition to overconfident mispricing, the market corrects only partially in the short run.

In a model with multiple securities, Daniel et al. (2001a) provide an analog to the CAPM when investors are overconfident. Security terminal cash flows satisfy a linear factor model, and each investor observes signals about the factors and the idiosyncratic component of security payoffs. Risk-averse investors form what they perceive to be mean-variance efficient portfolios. Overconfident individuals trade with risk-averse arbitrageurs who form rational beliefs. A security’s equilibrium expected return is linearly increasing in the security’s beta with the market, and the security’s current mispricing. Variables containing market price are proxies for the security’s misvaluation. For example, a fundamental/price ratio such as book/market is driven down when favorable news drives a stock up. Since there is overreaction, this is when the stock is overvalued. Thus, a high fundamental/price ratio
predicts high future returns. Aggregate value measures such as the market dividend yield or book/market positively predict future market returns.

A fundamental/price ratio (e.g., high book/market) tends to be high if either risk is high or if the market has overreacted to a highly adverse signal. In either case, price on average rises. Since high book/market reflects both mispricing and risk, whereas beta reflects only risk, book/market tends to be a better predictor of returns. These two sources of predictive power are unequal. Beta helps disentangle these cases, so beta and book/market are joint predictors of future returns.

However, when overconfidence becomes very strong, and if the proxy for the unconditional expected value (e.g., book value) is perfect, then the incremental ability of beta to predict future returns vanishes. The fundamental/price ratio dominates beta even though risk is priced. This is an extreme case, but it helps explain why empirical findings on the incremental effect of beta have been weak and inconsistent. The model also implies that in univariate regressions, beta should predict future returns. The model further describes the trade-offs in constructing optimal price-related proxies for misvaluation.

Daniel, Hirshleifer, and Subrahmanyam (2001b) extend the Daniel et al. (2001a) model to examine regressions of future returns on both book/market and HML loadings (Subsection II.A.2). They find that in an imperfectly rational model, either characteristics (e.g., book/market) or covariances (e.g., HML loadings) can be stronger predictors of future returns.

C. Dynamic Asset Pricing

Static models provide simple generalizations of the insights of the CAPM that can encompass the effect of risk as well as mispricing. However, a static approach has no hope of capturing the distinction between short-term continuation and long-term reversals. In both static and dynamic models, long-run reversal occurs when there is an overreaction to an impulse such as the arrival of good news. In a dynamic setting, short-run positive autocorrelation is consistent with long-run reversal so long as the process of overreaction and correction is sufficiently smooth. Such smoothness implies that when an impulse sets price rising, it will probably rise some more; that on average the last up-move to the peak of the impulse response function is not followed by a precipitous drop; and when the price is falling, it tends to fall some more. In contrast, a long-lag autocorrelation tends to associate positive returns during the overreaction process with negative returns arising during the correction process. The subsections that follow describe the effects of pure (independent) noise trading, mechanistic models based on correlated trading (positive feedback), the effects of mistaken beliefs, and the effects of alternative preferences.

C.1. Pure Noise Trading

Pure noise trading and positive feedback trading cause overreaction, and hence negative autocorrelations in long-run returns. When a stock rises too high, it needs to correct back down. Equivalently, this overreaction causes
excess volatility in returns. Furthermore, Campbell and Kyle (1993) show that overreaction can cause aggregate stock market value measures such as dividend yield to predict future market returns, so that contrarian investment strategies are on average profitable.

DeLong et al. (1990a) model the consequences of unpredictable random trades. Two securities pay identical, riskless dividends. The price of one asset is exogenously fixed. The other asset is risky because pure noise trades cause stochastic mispricing. Rational arbitrageurs with exogenous short time horizons limit their arbitrage trades for fear that the mispricing will get worse before it gets better. On average the risky asset trades at a discount, the risk premium demanded by the rational investors.

The noise-trading approach provides an explanation for the existence and behavior of closed-end fund discounts and their correlations with stock returns. According to DeLong et al. (1990a), noise traders buy and sell closed-end funds in a correlated fashion, causing discounts or premia relative to net asset value to fluctuate. The mispricing risk this creates makes these funds less attractive to rational investors, so on average there is a discount. This theory implies that fund discounts move together based on a systematic noise-trading factor; such comovement exists (e.g., Lee et al. (1991)). The theory also explains why such funds are created: to exploit optimistic noise traders.

Lee et al. (1991) suggest that shifts in fund discounts reflect shifts in noise-trader sentiment toward all small stocks. This is consistent with their evidence that narrowing of closed-end fund discounts is associated contemporaneously with high small stock returns. This implies that discounts predict small stock returns (see Section II). If discounts were a consequence of pure noise trading, they would be uncorrelated with future fundamentals, such as accounting performance. Swaminathan (1996) finds that at lags of greater than one year, high discounts predict both low future accounting profits and high future stock returns. This is consistent with fund investors overreacting to genuine information.

The comovement in small stock returns documented in Fama and French (1993) may come from correlated imperfectly rational trades (see Shleifer (2000), p. 20). The DeLong et al. (1990a) approach then suggests that small stocks, including closed-end fund shares, will earn high expected returns in compensation for their high mispricing risk. Alternatively, low market-value stocks may earn high returns because a stock’s low market value on average derives partly from its being undervalued (see, e.g., Daniel et al. (2001a)). The U.S. small firm effect has been weak or absent in the last 15 years, yet closed-end fund discounts remain.

C.2. Positive Feedback Trading

Positive feedback trading has several possible motivations, one being that investors form expectations of future prices by extrapolating trends (a topic covered in the next subsubsection). DeLong et al. (1990b) offer a model with a risky asset and risk-free cash, in which information arrives sequentially. The exogenous date 2 demand of the positive feedback traders is linearly
increasing in the preceding price trend. Foreseeing this, rational speculators buy into price trends, exaggerating trends, and overshooting. As a result there is excess volatility, and long-term negative autocorrelations in returns.

In Cutler et al. (1990), there are two types of imperfectly rational traders, positive feedback traders, and fundamental traders who ignore price and trade based upon a signal about the security’s payoff. Some fundamental traders observe this signal with a lag. This lag creates price trends which are profitably exploited by feedback traders. The gradual process of overshooting and correction induces both short-lag positive autocorrelation and long-lag negative autocorrelation.

More recent models with endogenous decisions have found things akin to pure noise trading—a limiting case of overconfidence, and positive feedback trading. But endogenously derived positive feedback is conditional and statistical, which seems more realistic than the older models. For reasons of both descriptiveness and predictiveness, explicit modeling of the psychology of investors is likely to supersede the mechanistic approach (except in otherwise-intractable applications).

C.3. Mistaken Beliefs

One explanation for return predictability is that investors set prices based on mistaken expectations. This subsection first considers dynamics when irrational individuals share the same biases (either overconfidence, or representativeness and conservatism). I then consider the interaction of multiple trader types with different biases.

C.3.1. The Dynamics of Biased Attribution and Overconfidence. Two recent papers provide models with a single risky security that reflect the fact that people learn about their own abilities in a biased, self-promoting fashion. In these models, investors do not know the precision of their private information signals, which reflects their information-gathering ability. They learn about their precision through time by observing whether later public news confirms or disconfirms their previous signal. The analyses assume the dynamic complement of overconfidence, biased self-attribution. When an investor receives confirming news, his confidence in his precision rises too much, and when there is disconfirming news, his confidence declines too little.

In Daniel et al. (1998), the impulse response function to a favorable initial shock, the private information signal, is hump-shaped. Price on average rises further as public information arrives, because confidence about the private signal on average grows. Eventually, however, accumulating evidence forces investors back to a more reasonable self-perception. This smooth hump-shaped impulse response implies positive short-lag and negative long-lag autocorrelation.

Shefrin and Statman (1994) analyze the general effect of mistaken beliefs on equilibrium prices in securities markets. They predict that when prices are inefficient, mispricing is related to a “beta correction”; it has not been obvious how to test this.
return autocorrelations. Daniel et al. (1998) also numerically simulate the correlation of a public information surprise (such as favorable accounting performance) with future returns with self-attribution bias. At short lags, this correlation is positive, but at long lags, the correlation can be negative (see Section II.A.4).

Gervais and Odean (2001) provide a model that accommodates analytical solution for the learning process under biased self-attribution. As traders become overconfident, trading volume and market return volatility increase. Since equity is in positive net supply, the model also predicts that trading volume will be higher after market rises than market falls, consistent with Statman and Thorley (1998).84

C.3.2. The Dynamics of Representativeness and Conservatism. Barberis, Shleifer, and Vishny (1998) offer an explanation for under- and overreactions based on a model in which actual earnings for a risky asset follow a random walk, but investors do not understand this. They mistakenly believe that the earnings process stochastically fluctuates between a regime with mean-reverting earnings, and a regime with expected earnings growth.

If recent earning changes reverse, investors erroneously believe the firm is in a mean-reverting state, and underreact to recent news, consistent with conservatism (Section I.A.4). If investors see a sequence of growing earnings, they tend to conclude (wrongly) that the firm is in a growth regime, and overextrapolate trends, which is arguably reminiscent of representativeness (Section I.A.3). Overreaction to a long enough trend implies subsequent low returns during the process of correction. Thus, there can be long-term overreaction and correction, implying negative long-lag return autocorrelation. Yet the average response to an initial impulse can be smooth, implying positive short lag autocorrelation. Similarly, the model can accommodate a positive short-term correlation between the asset return and an earnings change, and a negative long-term correlation. If sporadic events such as dividend initiations are viewed as isolated from earnings patterns, a single-event version of the model implies, under appropriate parameter values, underreaction.

Cross-sectional effects (such as a value effect) are simulated with earnings that are independently distributed across stocks. This implies a nearly risk-free arbitrage opportunity for a rational investor who buys and sells stocks based on return predictors. Such arbitrage would be risky in a setting where investors update their beliefs about systematic factors in earnings trends or reversals. The psychological literature on multiple cue learning (Section I.A) may provide guidance for such a model.

84 The implication of attribution/overconfidence models for whether there should be something akin to a disposition effect (holding winners, selling losers) is not obvious. When a stock is first becoming a winner, rational arbitrageurs who foresee further price rises should drive the price up even higher than the overconfident think is justified. This encourages the overconfident to sell, consistent with the disposition effect. However, for a stock that has been a winner for some time, the arbitrageurs will sell to the overconfident as the price peaks. Other recent models of momentum and reversal have similar opposing effects.
C.3.3. Interactions Among Traders with Different Biases. Hong and Stein (1999) analyze a market in which, as in Cutler et al. (1990), some traders react sluggishly, and others trade based on positive feedback. Each group of traders is risk averse, and is able to process only a subset of available information. Information about the liquidating dividend dribbles into the hands of different groups of news watchers. News watchers condition on their own private signals but ignore market prices, causing underreaction. Momentum traders, in contrast, condition on the cumulative price change over the last $k$ periods. Each trader takes a fixed position for a given number of periods. Momentum traders exploit the underreaction of news watchers by buying in response to price increases. This accelerates the reaction to news, but also causes overshooting. The smoothness of the overreaction process causes positive short-lag and negative long-lag autocorrelation. Slower information diffusion tends to launch a more powerful overreaction, leading to more negative long-lag autocorrelations.

C.3.4. Other Errors in the Dynamics of Beliefs. Although it is impossible to be comprehensive, I briefly mention some other approaches to the dynamics of beliefs. Shefrin (1997) discusses how base rate underweighting may shed light on the anomalous behavior of implied volatilities in options markets. Cecchetti, Lam, and Mark (1999) model the equity premium puzzle and related issues as arising from a combination of errors, including underestimation of the persistence of high- versus low-consumption growth regimes. They describe a rule-of-thumb calculation method that lead to such errors, but do not address whether other rule-of-thumb methods would imply the opposite error.

Informal arguments about money illusion affecting prices have been offered by several authors. Investors subject to money illusion may discount real cash flows at nominal interest rates, causing overdiscounting during high-inflation (growing inflation?) periods. They also may fail to take into account that higher inflation reduces the real value of a firm’s debt. Ritter and Warr (2001) provide evidence suggesting that inflation illusion contributed to the 1982 to 1999 bull market.

C.4. Alternative Preferences

Psychological evidence does not support the traditional assumption of time-additive expected utility. Theorists, often motivated more by puzzling securities price evidence than by psychological evidence, have offered models based upon alternative preferences. Alternative preference models can address the equity premium puzzle, the interest rate puzzle, and excess stock market volatility in at least two ways. First, by breaking the link between

---

85 Kurz (1997) describes his theory of endogenous uncertainty and rational belief equilibrium, which focuses on sets of beliefs that cannot be reliably contradicted by existing data. However, Bayesian updating has greater appeal as a theory of rational decisions.

86 Fisher (1928), Modigliani and Cohn (1979), Sharpe (1999), and Ritter and Warr (2001).
risk aversion and the intertemporal elasticity of substitution, a high equity risk premium (which demands high aversion to risk) can be reconciled with low interest rates (which demand reasonably high intertemporal elasticity of substitution). Second, by allowing risk aversion to vary stochastically, stock price volatility can be increased relative to consumption variability.

Several papers address the equity premium and risk-free rate puzzles applying habit-forming preferences.\(^{87}\) Constantinides (1990) shows that habit formation (Subsection I.E) reconciles a high equity premium with realistic consumption smoothness and growth, and moderate levels of risk aversion. Campbell and Cochrane (1999) and Chan and Kogan (2000) find that habit preferences that involve a Veblen-like concern for consumption of others imply stochastic risk aversion, which can reconcile a variety of facts about first and second moments of returns and consumption.

Several papers apply aspects of prospect theory and first-order risk averse preferences.\(^{88}\) Benartzi and Thaler (1995) consider investors who make a sequence of myopic single-period portfolio decisions. Consistent with loss-aversion, investors care about changes in wealth or consumption relative to a reference point that shifts from decision to decision, and their value function is kinked at the reference point. Investors, therefore, are highly averse to risks of short-term losses in stocks relative to bonds.

Shumway (1998) extends this approach to explain the cross section of expected returns as well as the market expected return. Consistent with prospect theory, he assumes a modified power utility function that implies risk aversion over gains and risk seeking over losses. The reference point is a zero market return. In consequence, small market returns cause relatively large changes in the stochastic discount factor. In equilibrium, stock prices are a linear function of the stock's upside beta and its downside beta.

Empirically, Shumway (1998) finds that the model does quite well in fitting both the equity premium puzzle and the cross section of security returns. He suggests that the high premium on equity results from loss aversion, which causes marginal utility to vary more with slightly negative market returns. This tends to magnify the effect of stocks' downside risk relative to that of bonds.

Barberis et al. (2001) offer a model based upon a combination of loss aversion, and the “house money” effect of Thaler and Johnson (1990), the tendency for individuals who have experienced recent gains to be less averse to risky gambles. To capture loss aversion, they assume a piecewise linear value function that is steeper among losses than among gains relative to the reference point. After the good news of a high dividend, individuals become more risk tolerant. Stochastic variation in risk aversion increases the vola-


tility of returns relative to dividends. These fluctuations in risk aversion tend to reverse, causing predictability in stock returns. The high return variability raises the equity risk premium even without high aversion to consumption risk, and is therefore consistent with a reasonably low risk-free rate.

Barberis and Huang (2000), like Shumway (1998), examine the dynamics of loss aversion with many risky securities. Barberis and Huang (2000) consider two kinds of mental accounting. Under individual stock accounting, investors care about total consumption, but are also loss averse over individual stock movements. In the other, portfolio accounting, individuals are loss averse with respect to movements in their total stock portfolio.

Investors are also subject to the house money effect. Using plausible parameter values, under individual stock accounting, the typical individual stock has a high expected excess return, and its returns are variable relative to dividend variability. The cross section of returns is predictable using measures of size, value, and whether the firm was a winner or loser over the last three years. The model implies an even higher equity premium than Barberis et al. (2001) because investors are loss-averse with respect to the residual risk of individual stock movements.

In a broadly similar spirit, Epstein and Zin (1993) examine a first-order risk averse setting and report that the case of disappointment averse preferences fit the data well (see also Epstein and Zin (1990)). Bekaert et al. (1997a) find that first-order risk aversion can explain predictability in U.S. and Japan equity, bond, and foreign exchange markets better than the expected utility model, but not enough to match the data. Ang et al. (2000) find that a high U.S. equity premium is consistent with reasonable parameters of disappointment averse preferences.

A rather different approach from applications of loss-averse or first-order risk averse preferences focuses on aversion to ambiguity (Subsection I.C.1) and a consequent taste for robustness. A robust decision rule is one that does well in the face of model uncertainty when Nature chooses the most adverse possible model in response to the individual’s choice. Tornell (2000) provides a model based upon agents who choose robust forecasting techniques to explain high equity returns, predictability, and excess volatility.

Even slight stochastic shifts in preferences can substantially increase the volatility of stock prices relative to the variability of consumption (Allen and Gale (1994), Kraus and Sagi (2000), and Mehra and Sah (2000)). The psychological evidence that visceral factors affect decisions are consistent with such variability.

C.5. Evolving Populations

A promising field for exploration uses evolutionary simulation of the interactions of agents in financial markets. In the last five years, physicists have begun to do research on financial markets, some calling their field econo-

---

89 See, for example, Hansen, Sargent, and Tallarini (1999) and Maenhout (2000).
physics (see Farmer (1999)). Some of the recent models by physicists make such radical mechanistic assumptions about investor behavior and market structure that the resulting insights seem unlikely to generalize. Fortunately, a very promising strand of evolutionary literature explores the populations of traders who are imperfectly rational but do learn and make endogenous decisions. Freed from the constraints of analytical tractability, modelers are able to explore a wider space of economic settings.

An evolutionary approach could address the argument that even though individuals are imperfectly rational, as they learn from their trading outcomes, the system will progress toward the fully rational equilibrium rapidly. I conjecture that a simple tropism among traders towards actions that generate higher investment profits will not converge to the ICAPM quickly. Even with a long history of evidence, it is hard for a trader to figure out whether a trading strategy has done well after adjusting for risk unless he understands risk, and the income and substitution effects of Merton ICAPM hedging are, I conjecture, too subtle for most individual or even sophisticated institutional investors.

Some brief recent surveys of the computational field include Farmer (1999), Farmer and Lo (1999), and Lebaron (2000a). Some recent findings are that long-horizon investors frequently do not drive shorter-horizon investors out of financial markets, and that populations of long- and short-horizon agents can create patterns of volatility and volume similar to actual empirical patterns (Lebaron (2000b, 2000c)).

D. Empirically Distinguishing Pricing Theories

The effects described in different psychological pricing theories need not be mutually exclusive, but it is useful to examine how their predictions differ. My focus is on value, momentum, and event-based effects.

D.1. Distinguishing Explanations for Size and Value Effects

Several past authors have pointed out that long-run overreaction will induce cross-sectional value effects. Two recent models derive cross-sectional value and size effects when securities are subject to systematic and idiosyncratic influences (Barberis and Huang (2000) and Daniel et al. (2001a)).

Daniel et al. (2001a) provides no help in explaining the equity premium puzzle. The Barberis and Huang (2000) theory does address the equity premium and associated puzzles because people who do individual-stock accounting are averse to the high residual risk of stocks. Thus, a further implication of Barberis and Huang (2000) is that residual risk is cross-sectionally priced (Brennan (2001)). Furthermore, in contrast with the Merton (1987) limited participation theory, the Barberis and Huang (2000) theory seems to imply that, ceteris paribus, greater participation by individual investors will increase the premium for residual risk.

Daniel et al. (2001a) offer further implications, largely untested, concerning the cross-sectional dispersion in fundamental/price ratios, and the ability of current volume to predict future return volatility. Another implication
is that as confidence exogenously varies over time, the dispersion in security fundamental/price ratios varies together with the ability of such ratios to predict future returns. Cohen, Polk, and Vuolteenaho (2000) confirm such a relationship between the book/market “value spread” and the profitability of value trading strategies.

In Daniel et al. (2001a), mispricing is present in a small number of factors. The importance of idiosyncratic effects in the Barberis and Huang (2000) theory suggests that rational arbitrageurs should take strong contrarian positions and earn large expected profits. More broadly, the strong flow of wealth in the Barberis and Huang (2000) theory suggests that mispricing effects should be more transitory than in Daniel et al. (2001a). Barberis and Huang (2000) also point out that the rise of mutual and pension fund stock investments should have led to less individual stock accounting, and is, therefore, consistent with a weakening in size and value effects.

There is psychological evidence that overconfidence is strongest when information signals are less precise and when feedback is inconclusive (e.g., Einhorn (1980) and Griffin and Tversky (1992)). Thus, Daniel et al. (2001a) predicts that fundamental/price ratios should forecast risk-adjusted returns more strongly for businesses that are hard to value (e.g., R&D-intensive firms comprised largely of intangible assets). Chan, Lakonishok, and Sougiannis (1999) subsequently report evidence consistent with such a pattern.

Neither Barberis and Huang (2000) nor Daniel et al. (2001a) capture momentum. The absence of a unified model that directly captures the two most conspicuous cross-sectional effects, value and momentum, is an obvious gap in the literature. The results of Daniel et al. (1998) and of Barberis et al. (1998) suggest that unified explanations may be possible based upon either overconfidence, or upon misperceptions of regime-shifting.

D.2. Distinguishing Explanations for Postevent Continuation

The Daniel et al. (1998) analysis of postevent continuation differs from the Barberis et al. (1998) model in predicting continuation only for selective events taken by a party such as management or an analyst in response to market mispricing. The support for this from one type of event (see Subsection II.A.4 at footnote 62) is intriguing. Event studies on other low-discretion events (such as regulatory announcements, input supply shocks, or output demand shocks) provide an attractive direction for further testing.

The Barberis et al. (1998) model is based upon public information. The Daniel et al. (1998) model implies negative long-run return autocorrelation associated with private information arrival, and positive long-run return autocorrelation associated with public signals. This is consistent with evidence of Daniel and Titman (2000). Daniel et al. (1998) further predict that postevent continuation will be strongest in stocks about which investors have poor information (often illiquid or smaller stocks). Daniel et al. (1998) also offer several untested predictions about the occurrence of and price patterns around corporate events, and about volatility at the time of private versus public signals.
D.3. Distinguishing Explanations for Momentum and Reversal

Analytically, the three recent models of how mistaken beliefs cause momentum and reversals (Barberis et al. (1998), Daniel et al. (1998), and Hong and Stein (1999)) all generate an impulse response function to a new information signal in which there is a gradual rise in the average reaction to a positive signal and a gradual average process of correction.

In all these models, the misperceptions that drive momentum are also the drivers of long-term reversal. These models therefore imply that if there is some market segmentation, then those sets of assets with largest momentum effects should also have the largest reversal effects; international testing would be of interest. It is interesting that much of the empirical evidence of return predictability, including both momentum and reversal, is stronger in small firms (see Fama (1998) and Loughran and Ritter (2000)). Barberis et al. (1998) and Daniel et al. (1998), but not Hong and Stein (1999), further imply that those countries or sets of securities with strongest momentum effects should also have strongest post-earnings-announcement drift.

More generally, greater uncertainty about a set of stocks, and a lack of accurate feedback about their fundamentals, leaves more room for psychological biases. At the extreme, it is relatively hard to misperceive an asset that is nearly risk-free. Thus, the misvaluation effects of almost any mistaken-beliefs model should be strongest among firms about which there is high uncertainty/poor information (cash flow variance is one possible proxy). Furthermore, in Daniel et al. (1998) and Hong and Stein (1999), greater information asymmetry strengthens the predicted effects; the adverse selection component of the bid-ask spread is a possible proxy. Barberis et al. (1998) does not have implications based on information asymmetry. Firm size, analyst following, and dispersion in analyst forecasts are potential proxies for information asymmetry, but they also may proxy for mere uncertainty. Thus, evidence that small firms (internationally) and firms with low analyst following have greater momentum is consistent with, but does not sharply distinguish, the three models.

Barberis et al. (1998) predict overreaction to trends, which can also occur in Daniel et al. (1998), but it is not obvious that the Daniel et al. (1998) implication extends to zero net supply securities. Thus, the evidence of Poteshman (2001) of daily underreaction and multiple-day overreaction of option prices to shifts in volatility supports Barberis et al. (1998) (at a very different time horizon).

Bloomfield and Hales (2001) directly test the Barberis et al. (1998) theory that people misperceive random walks to be shifts between continuation and reversal regimes by examining predictions by MBA-student experimental subjects. Consistent with Barberis et al. (1998), subjects overreacted to changes preceded by sequences of continuations, and underreacted to changes preceded by many reversals. However, people on average tended to expect continuation, whereas a perceived tilt toward reversal is needed to obtain postearnings announcement drift and postevent return continuation.
Another testing approach is to find data sets in which the trades of irrational traders versus rational arbitrageurs can be identified. Coval and Shumway (2000) analyze a rich database to describe how the positions of futures market-makers change following recent trading success. Another suggestion has been to view market orders as irrational and limit orders as rational (Hvidkjaer 2000). However, it does not seem clear why this would be the case based upon these theories, and empirically it is limit order traders who lose money in the short run (Chordia, Roll, and Subrahmanyam 2000)). Further progress on these issues calls for explicit modeling of psychology and microstructure.

IV. Conclusion

Man is neither infinite in faculties, nor in apprehension like a god. Nor is human fallibility shed at the doorstep of the stock exchange. Psychology-based asset-pricing theory has promise of capturing this reality, though at this point we are at an early stage.

Financial economists have grown more receptive to entertaining psychological explanations. One sign of this is the popularity of utility functions that seem to violate time consistency or the rationality axioms of expected utility in recent literature on the equity premium and risk-free rate puzzles. Some of these preferences could be endogenized as reduced form summaries of rational settings with market frictions, but this does not seem to be a high research priority even among fans of the full-rationality approach.

In Section I, I tried to give some hint of the wealth of psychological findings, many utterly unexploited, that can inform financial modeling. Subsection III.D offered hints for empirical work to distinguish alternative psychology-based pricing theories. I now mention a few other possible theoretical and empirical directions.

1. So far few psychology-based asset-pricing models allow for both risk aversion and multiple risky securities. It will be useful to explore the dynamics of mistaken beliefs when there is a cross section of securities, to address such issues as volume as a predictor of returns, and the effects of different rates of overreaction and correction for factors and residuals.

2. Pricing models based upon loss- and disappointment-aversion can be viewed as reflecting a concern about future feelings. But more directly, the effect of currently experienced emotions on current prices merits analysis.

---

90 Conjecturally, Daniel, Hirshleifer, and Subrahmanyam (1998) implies that rational arbitrageurs buy after good news (foreseeing further overreaction). Overconfident traders sell (because the arbs drive prices up even higher than justified based on current overconfident beliefs). Some period of time after the favorable impulse, the arbs tend to sell out to the overconfident, and to go short. Anticipation by arbs of overreaction should generate similar trading patterns in Barberis et al. (1998); in the Hong and Stein (1999) setting, the behavior of irrational traders is more complex.
3. It is often not obvious how to translate preexisting evidence from psychological experiments into assumptions about investors in real financial settings. Routine experimental testing of the assumptions and conclusions of asset-pricing theories is needed to guide modeling.
4. We lack a quantified set of capital budgeting and risk-management procedures that reflect mispricing and are ready for practitioners to apply (but see Stein (1996)).
5. The great missing chapter in asset-pricing theory, I believe, is a model of the social process by which people form and transmit ideas about markets and securities. In addition to studying what influences individuals’ valuations, an appealing direction is to study how attention is focused on certain groups of stocks, and the effects of resulting swings in participation. A different empirical direction is to analyze the specific content of widespread, erroneous investor theories to identify ways of predicting returns. Robert Shiller has discussed and documented investor theories, belief transmission, and effects on pricing (e.g., Shiller (1984, 1990, 2000c); see also the analysis of DeMarzo, Vayanos, and Zwiebel (2000)). This research has blazed a path upon which further work will follow.

My list of further directions is necessarily idiosyncratic. In an area that is just coming of age, many new prospects are open. This is an exciting time for the field of asset pricing.

REFERENCES

Abel, A. B., 1990, Asset prices under habit formation and catching up with the Joneses, American Economic Review 80, 38–42.
Asch, S., 1956, Studies of independence and conformity: A minority of one against a unanimous majority, *Psychological Monographs* 70 Whole No. 416.
Barber, B., and T. Odean, 1999, Online investors: Do the slow die first? University of California–Davis.
Barber, B., and T. Odean, 2000a, Boys will be boys: Gender, overconfidence, and common stock investment, *Quarterly Journal of Economics*, forthcoming.
Barberis, N., and M. Huang, 2000, Mental accounting, loss aversion, and individual stock returns, University of Chicago.


Bloomfield, R., and J. Hales, 2001, Predicting the next step of a random walk: Experimental evidence of regime-shifting biases, Johnson School, Cornell University.


Cao, H. H., and D. Hirshleifer, 2000, Conversation, learning and informational cascades, Working paper, Ohio State University.


Chan, Y. L., and L. Kogan, 2000, Catching up with the Joneses: Heterogeneous preferences and the dynamics of asset prices, Wharton School #14-00.


Chordia, T., R. Roll, and A. Subrahmanyam, 2000, Order imbalance, liquidity, and market returns, Working paper, UCLA.


Fair, R., 2000, Events that shook the market, Working paper, Yale University.


Harmon-Jones, E., and J. Mills, 1999, An introduction to cognitive dissonance theory and an overview of current perspectives on the theory, in Eddie Harmon-Jones and Judson Mills,


Kirby, K., and R. J. Herrnstein, 1995, Preference reversals due to myopic discounting of delayed reward, Psychological Science 6, 83–89.


Kothari, S., 2000, Capital market research in accounting, Working paper, MIT.


LeBaron, B., 2000b, Evolution and time horizons in an agent-based stock market, Working paper, Brandeis University and NBER.
LeBaron, B., 2000c, Evolving long run investors in a short run world, Working paper, Brandeis University and NBER.
Lewis, K. K., 1999, Trying to explain home bias in equities and consumption, *Journal of Economic Literature* 37, 571–608.
Lichtenstein, S., and B. Fischhoff, 1977, Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance* 20, 159–183.


Longstaff, F. A., P. Santa-Clara, and E. Schwartz, 1999, Throwing away a billion dollars: The cost of suboptimal exercise strategies in the swaptions market, Anderson School Working paper #8-00, UCLA.


Nisbett, R. E., and T. D. Wilson, 1977b, Telling more than we can know: Verbal reports on mental processes, Psychological Review 84, 231–259.


Trivers, R., 1985, Social Evolution (Benjamin/Cummings, Menlo Park).


Yariv, L., 2001, I’ll see it when I believe it—A simple model of confirmatory bias and cognitive dissonance, Mimeo, Harvard University.