Experimental Research in Financial Accounting

Robert Libby

Robert Bloomfield

Mark W. Nelson

Johnson Graduate School of Management
Cornell University
Ithaca NY 14853

February 23, 2001

Prepared for the *Accounting, Organizations and Society* 25th Anniversary Conference, Oxford University, July 2000. We thank Ron King, Lisa Koonce, Laureen Maines, Greg Waymire, the participants at the *AOS* 25th Anniversary Conference and the Emory Behavioral Financial Accounting Research Conference for their comments and suggestions, and Bernadine Low for her assistance.

Experimental Research in Financial Accounting

1. Introduction

Financial accounting research is a broad field that examines financial communication between managers, auditors, information intermediaries, and investors, as well as the effects of regulatory regimes on that process. Much of this literature focuses on managers' and auditors' reporting decisions and their relationships to analysts' forecasts and value estimates, investors' trading decisions, and resulting market prices. This clear focus on judgment and decision making led to the large number of experimental financial accounting studies published in major accounting journals in the 1960s and 1970s.

Serious criticisms of this early research (e.g., Gonedes and Dopuch 1974) turned experimentalists' focus away from financial accounting issues in the 1980s and early 1990s. As discussed by Maines (1995) and Berg, Dickhaut, and McCabe (1995), major elements of these criticisms were: (1) the irrelevance of individual behavior in market settings, in which competitive forces will eliminate individual "errors," (2) poor matching of research methods to research questions, (3) the lack of psychological or economic theory to predict effects and specify the mechanisms through which they occur, and (4) failure to capture relevant aspects of the decisions of interest, in particular, decision maker attributes and institutional features.

Beginning in the mid-1990s, there was a resurgence of experimental research addressing an even broader spectrum of financial accounting issues. This paper presents our view of how this new literature has addressed prior criticisms, and how it can continue to shed light on financial accounting questions. We argue that significant evidence of capital market inefficiency has renewed interest in how individuals make key accounting-related decisions and how these decisions affect market prices. Recent studies take advantage of the experimentalist's comparative advantage at disentangling variables that are confounded in natural settings and measuring intervening processes to draw strong causal inferences. Theories combining psychology and economics have allowed experimentalists to specify more clearly the

mechanisms affecting individual and market behavior. Finally, most of the new studies focus on issues of clear relevance to financial accounting, particularly the effects of decision-maker knowledge and motivation, the complex information environment, regulation, and strategic interaction.

This paper is aimed primarily at those who plan to conduct financial accounting experiments, and secondarily at other financial accountants who are interested in what can be learned from experimental studies. Our primary goal is to use recent experimental studies of financial accounting to illustrate our view of how such experiments can be conducted successfully. The core of our view is that *successful financial accounting experiments use the comparative advantages of the experimental approach to determine how, when and (ultimately) why important features of financial accounting settings influence behavior.* By elaborating on this view, we hope to increase the impact of future experiments and help the new literature avoid the mistakes and fate of the earlier literature. We do not provide an exhaustive review of the literature, nor do we provide detailed critiques of particular studies. Instead, we focus on how particular examples illustrate successful use of experiments to address important financial accounting issues. Our examples include and integrate experiments that fall into both the "behavioral" and "experimental economics" literatures in accounting.[1] Although these literatures evolved from different traditions, we see them as essentially similar—both use experiments to shed light on financial accounting issues, and therefore both present similar opportunities and challenges to researchers. Naturally, our review is also deeply affected by our own biases and the financial accounting issues that we have been addressing in our own recent research.

In section 2, we describe in more detail how changes in views of market efficiency, reliance on the experimentalist's comparative advantage, new theories, and a focus on key institutional features have allowed recent experiments in financial accounting to overcome the

[1] See Haynes and Kachelmeier (1998) and Moser (1998) for recent discussions of the integration of the behavioral and economic approaches to experimentation.

criticisms of the earlier literature. In section 3, we describe how specific streams of experimental financial accounting research have addressed questions about financial communication between managers, auditors, information intermediaries, and investors, and indicate how future research can extend those streams. We focus particularly on (1) how managers and auditors report information, (2) how users of financial information interpret those reports, (3) how individual decisions affect market behavior, and (4) how strategic interactions between information reporters and users can affect market outcomes. While we address studies of auditors in their financial reporting role, to limit the scope of the review, we do not address issues related to the demand for and conduct of auditing. We also do not address studies of creditors' decisions, which have received little attention in recent financial accounting experiments.

In section 4, we discuss how experiments can be designed to be both effective and efficient. We use the "predictive validity framework" (Runkel and McGrath 1972; Libby 1981) to structure our discussion of maximizing effectiveness through careful hypothesis development and research design. Our discussion of efficiency focuses on the consumption of scarce resources, such as subjects and compensation to those subjects. We conclude in section 5 with a brief summary of our main points.

2. Factors Affecting the Supply and Demand for Experimental Financial Accounting Research

In this section we examine four interdependent factors that have mitigated concerns raised about the earlier experimental literature and promoted recent progress in experimental financial accounting research: changing views of market efficiency, recognition of the strengths and weaknesses of experimental methods in addressing financial accounting questions, the availability of new theoretical bases for the research, and a more detailed view of the institutional features of financial accounting settings. We discuss each of these factors in turn.

2.1. Changing Views of Market Efficiency

Much of financial accounting research in the 1960s implicitly assumed that some investors' failure to adjust fully for the effects of accounting method choices would affect

allocation of resources in the economy and disadvantage these less sophisticated investors in their exchanges with more sophisticated investors (see Maines 1995 for a review). A series of papers in finance (particularly Fama 1970) persuaded many accounting researchers that if just a small fraction of investors are sophisticated enough to respond appropriately to accounting information, they will compete among themselves to set security prices equal to their expected values. As a result, the market becomes a "fair game" in which even unsophisticated investors are protected by the informational efficiency of prices.[2] This research led Gonedes and Dopuch (1974), among others, to argue that experimental research on individual behavior could have only limited importance for financial accounting.

In the late 1980s and 1990s, however, numerous studies reported market inefficiencies.[3] One line of research provides direct support for the assumptions underlying early financial accounting research: accounting policies affect pricing, even when they have no true economic effects (e.g., Hand 1990; Vincent 1997; Andrade 1999; Sloan 1996). Another line of research indicates more generally that fundamental analysis of public financial statement information can lead to higher stock returns (e.g., Ou and Penman 1989; Lee, Myers, and Swaminathan 1999; Frankel and Lee 1998). A third line of research suggests that even sell-side analysts—generally recognized as among the most sophisticated users of financial statements—are predictably biased (DeBondt and Thaler 1990; La Porta 1996; Dechow and Sloan 1997).

The best-known lines of efficiency research focus on momentum in earnings and prices. A voluminous literature on post-earnings-announcement drift shows that markets underreact to large earnings surprises (Foster, Olsen, and Shevlin 1984; Bernard and Thomas 1989, 1990; Bhushan 1994; Ball and Bartov 1996; Brown and Han 2000). Another literature, primarily published in finance journals, shows that after adjusting for risk, stock returns are positively autocorrelated over periods of several months (e.g., Chan, Jegadeesh, and Lakonishok 1996), but negatively autocorrelated over periods of several years (DeBondt and Thaler 1985, 1987).

[2] Watts and Zimmerman (1986) also provided particularly influential arguments.
[3] See Fama (1998), Kothari (2000), and Thaler (1999) for more comprehensive reviews of this literature.

The literature on market inefficiency is controversial, and many of the papers alleging inefficiency have been criticized on methodological grounds (see Ball 1992; Fama 1998; and Kothari 2000). Nevertheless, many researchers now doubt whether markets satisfy the requirements of the semi-strong form of the efficient markets hypothesis (that markets respond efficiently to all publicly-available information), or even the weak form (that markets respond efficiently to information contained in past market prices). Even some of the most skeptical seem to be convinced that post-earnings-announcement drift is not simply an artifact of research design (Ball 1992). Recent research on efficiency has also led theorists to examine how the assumptions underlying the efficient markets hypothesis might be relaxed to account for archival results. (We discuss these models more in section 2.3). As a result, experimental researchers can more easily argue that individual behavior can be an important element in determining market behavior, even in the presence of competitive forces.

2.2. The Comparative Advantage of Financial Accounting Experiments

Earlier financial accounting experiments typically sought to determine whether specific accounting policy choices would affect investors' decisions. Answers to such research questions call for estimates of the magnitude of an effect (or error) by representative actors in representative circumstances, a task ill suited to experiments. Such a task is more appropriate for archival-empirical research, which examines large representative samples of naturally occurring phenomena.

More recent experimental research strives to use experimentalists' comparative advantage to focus on disentangling the effects of variables that are confounded in natural settings and determining under what circumstances and through which processes specific phenomena arise. Experiments are well suited to this task because they construct their own research setting. In a constructed research setting, one can *manipulate* the independent variables, *control* for other potentially influential variables by holding them constant or through randomization, and *measure* the intervening processes (such as information search or the path

players take to equilibrium outcomes in strategic settings) and mental states (such as knowledge, beliefs, or confidence) that affect final outcomes. This allows an experimentalist to disentangle the effects of variables that are confounded in the environment to draw strong causal inferences, and to test the effects of conditions that do not yet exist or do not exist in sufficient quantity in the natural environment (Libby and Luft 1993). Experiments testing how and why (rather than whether or not) financial accounting phenomena occur can be based on theories of psychological, economic or institutional processes. We discuss these theories next.

2.3. Theoretical Advances in Psychology, Finance, and Economics

Earlier experimental research was criticized for the lack of psychological or economic theory that specified the mechanisms through which effects of accounting disclosures would occur. Recent experiments in financial accounting can rely on well-developed psychological theories of judgment and decision making[4] that were in their infancy when the studies reviewed by Gonedes and Dopuch (1974) were conducted. Recent research can also rely on economic models that describe more carefully when and how equilibrium outcomes arise.

The major idea underlying much research on judgment and decision making is that decision makers are boundedly rational (Simon 1957). Decision-makers often have limited information on which to base their judgments and decisions, limited ability to retain and retrieve that information from memory, limited ability to process and use that information, and limited insight into their own decision processes and future preferences. Studies over the last 25 years have focused on how various attributes of human cognition determine exactly what humans do well and what they do poorly. A number of their findings have influenced recent thinking in financial accounting and the study of financial markets.

Many decision-making studies emphasize the role of heuristics (Tversky and Kahneman 1974). Heuristics are simplified decision rules developed to deal with complex situations. These

[4] Syntheses of the key constructs or ideas that drive psychological theories of judgment and decision making have been provided by Carroll and Johnson (1990), Hogarth (1993), Bazerman (1998), and others.

heuristics are efficient and often work well. But in some circumstances they may lead to systematic biases such as over- and under-confidence in judgment (Griffin and Tversky 1992) and misperceptions of the covariation between signals and events (Lipe 1991), which can systematically affect the manner in which individuals react to financial accounting information and the manner in which that information is impounded in prices. Learning to overcome biases is difficult because of the uncertainty and poor feedback inherent in complex environments. Often what we learn from experience is not valid (Einhorn 1980).

The importance of (imperfect) storage and retrieval of information from memory has also been recognized in recent financial accounting experiments. Some of these studies rely on models of memory organization (e.g., Smith and Medin 1981) that indicate how knowledgeable decision makers efficiently organize and retrieve data. Other studies recognize that memory for events is influenced by factors that are normatively relevant, such as their frequency of occurrence, and factors that are normatively irrelevant, such as primacy, recency, and contrast effects (e.g., Hogarth and Einhorn 1992). Still others recognize that the limited capacity of working memory affects our ability to consider multiple factors in making a judgment or choice. Consequently, even normatively relevant factors that decision makers are aware of often times have limited influence on their judgments and decisions.

Recent research in accounting and finance also relies on psychological models of risk (e.g., Kahneman and Tversky 1979) and ambiguity (e.g., Einhorn and Hogarth 1986) that characterize individuals' responses to risk and reward in ways that deviate from standard expected utility theory.[5] This more recent psychology literature provides greater ability to predict under what circumstances behavior will be more or less likely to differ from the predictions of standard economic theory (e.g., in earnings predictions versus trading behavior, in different information environments). A large literature on social psychology could also be used to understand interaction between participants in financial accounting settings. For example,

[5] See Hodder, Koonce and McAnally (2001) for further discussion of risk in financial accounting settings.

research related to accountability (e.g., Tetlock 1992), motivated reasoning (e.g., Kunda 1990) and group decision processes (e.g., Yetton and Bottger 1982) has significantly influenced auditing studies.

Other financial accounting studies use advances in financial economics to test the assertion that biased traders will be driven out of the market through systematic trading losses. Some of these models focus on how biases might influence market outcomes. For example, Barberis, Shleifer, and Vishny (1998) use psychological models of how people perceive random-walk sequences in a model with a representative investor. Daniel, Hirshleifer, and Subrahmanyam (1998), Gervais and Odean (1997) and Odean (1998) incorporate overconfidence into trading models. Other models focus on forces that keep unbiased traders from exploiting price errors. For example, De Long, Shleifer, Summers, and Waldmann (1991) show that traders who respond irrationally to irrelevant information ("sentiment") create enough noise in prices to keep rational traders from exploiting the resulting price errors. Fischer and Verrecchia (1999) and Kyle and Wang (1997) show that overconfidence, although irrational, can actually give traders higher payoffs than their rational compatriots. These results make it difficult to argue that some form of natural selection will eliminate irrational traders in dynamic equilibria, and provide accounting researchers with specific models of how and when individual biases might influence market prices.

Experiments focusing on game theoretic models of financial accounting settings can now rely on new economic models that move beyond the traditional equilibrium view. Rather than simply identifying an equilibrium and assuming that it will occur, many economists have examined in detail what assumptions about rationality must be satisfied for equilibria to have predictive power (Bernheim 1984; Pearce 1984; Tan and Werlang 1988). Other models have examined the process by which equilibria are achieved, using either psychological theories based on behaviorism (Herrnstein and Vaughn 1980) or evolutionary theories of natural selection (Maynard Smith 1982). In a similar vein, Gode and Sunder (1993, 1997) used such ideas to show that "zero-intelligence" traders, who do nothing more than avoid obviously horrible strategies,

can achieve efficient security allocations in some markets. By focusing on processes by which equilibria are achieved, these studies provide indications of when equilibria will and will not predict behavior in financial accounting settings.

2.4. Key Institutional Features of Financial Accounting Settings

Most early experimental studies in financial accounting took relatively narrow views of financial accounting institutions. They typically focused on the set of rules governing how accounting information could be reported in financial statements, implicitly assuming that reporting choices (and interpretations of those choices) were made neutrally, rather than being influenced by the incentives of a strategic manager or auditor. Early studies also implicitly assumed that responses to financial accounting information would be independent of the expertise or incentives of the user, and that interactions among users and reporters would not alter outcomes.

Consistent with the advice of Libby and Luft (1993), recent experimental research in financial accounting has considered institutional features more broadly, and has also focused on the interaction between individual and environmental characteristics. Two key individual characteristics are the *knowledge* and *motivation* of information reporters and users. These determine the parties' goals, and how they use financial accounting to achieve those goals. Key environmental characteristics include the *complex regulations* governing reporting, the *existence of financial markets,* and the *strategic interactions* between reporters and users, as well as between different sets of users. Regulations determine the set of choices open to managers and auditors, and may also determine the results of those actions (e.g., lawsuit outcomes). Financial markets affect how individual decisions result in aggregate market outcomes, such as stock prices, liquidity and trading volume, and may also determine wealth transfers among different sets of investors. Strategic interactions capture the intertwining of the incentives and actions of the many parties to financial accounting decisions. Financial accounting settings include managers, auditors, investors and information intermediaries (analysts and the press) who may

all interact strategically. Managers and auditors negotiate to determine the contents of the financial statement and audit report. Investors draw inferences about managers' and analysts' information and incentives from observing reports. Managers may choose reports in an attempt to "fool" investors, but the investors may be able to anticipate these attempts.[6]

Focusing explicitly on individual and environmental characteristics allows experimental researchers to shed light on how and when experimental results will generalize to target settings, and also indicate how variations in these institutions will alter behavior. In this way, an institutional focus helps researchers to exploit the comparative advantage of experimentation. In the next section, we describe how specific streams of experimental financial accounting research have done so, and indicate how future research could extend those streams.

3. Key Financial Accounting Questions and Experimental Evidence

The goals of the literature that we review are similar to those of the broader financial accounting literature: to increase our understanding of the financial reporting process and its effects. While all of the studies that we examine share the same general goal, they focus on different elements of the interactions of boundedly rational managers, auditors, information intermediaries, and investors. These differences in emphasis led us to divide the studies into four related categories described by the following questions:

1. How do managers' and auditors' incentives and financial accounting regulations determine how they report events?
2. How do knowledge of accounting regulations, managers' incentives, and the information content of accounting reports affect users' (investors and information intermediaries) interpretations of accounting reports?
3. How do individual responses to information affect market-level phenomena?

[6] Financial accounting information is also used for contracting and stewardship purposes, but that has not been the focus of significant experimental research.

4. How do strategic interactions between reporters and users of information affect reporting and market outcomes?

We focus primarily on papers published since the publication of Maines's (1995) review of this literature.

3.1. How do managers' and auditors' incentives and financial accounting regulations determine how they report events?

Reporting performance is fundamental to financial accounting. Discretion provided by financial accounting regulations, coupled with the inherent subjectivity of much accounting measurement, allows managers some flexibility to opportunistically report or manage earnings. Consequently, much archival and experimental research has focused on this area.

Archival studies typically examine opportunistic reporting by identifying whether earnings or accruals differ from expectation in a manner favored by managers' incentives (see Healy and Wahlen 1999 for a review). While these studies have demonstrated numerous instances of apparent earnings management, their conclusions are sometimes criticized because of methodological difficulties, including poor incentive proxies, misstated discretionary accruals models, or potential omitted variables such as operating choices that have non-earnings-management rationales but that affect discretionary accruals (Dechow, Sloan, and Sweeney 1995; Bernard and Skinner 1996). Also, archival studies of earnings management focus on post-audit financial statements that are a joint product of the negotiations between managers and auditors, which makes it difficult to distinguish the separate contributions of managers and auditors to earnings management or to determine how managers' and auditors' separate incentives influence their reporting and attesting behavior (Nelson, Elliott, and Tarpley 2000).

Experimental studies avoid these problems by manipulating incentives and assessing treatment effects rather than attempting to measure unexpected accruals, and by holding constant task characteristics that create potential omitted variables problems. Experiments can examine managers' and auditors' judgments separately, but can also examine auditor-client interactions.

These characteristics of experimental work have led to a growing experimental literature that complements the archival work in this area.

The largest group of experimental earnings-management studies focuses on *auditors'* incentives and the circumstances under which they allow managers to take aggressive accounting positions. Consistent with the general auditing literature (e.g., Kinney and Martin 1994), results indicate that auditors reduce the aggressiveness of financial reports. For example, Hirst (1994) provides evidence that auditors consider management competence and objectivity when evaluating management-provided evidence. Phillips (1999) demonstrates that, after auditors receive evidence of aggressive reporting in high-risk accounts, they are more likely to attend to it elsewhere, even in accounts they typically consider to be of low risk. Kinney and Nelson (1996) demonstrate a circumstance in which auditors make audit-reporting judgments that are as conservative as thought appropriate by even those investors who are evaluating the audit report in the presence of negative outcome information.

However, other studies indicate that auditors are more likely to allow their clients to take aggressive accounting positions when the relevant evidence or precedents offer more room for interpretation. For example, Nelson and Kinney (1997) provide evidence that auditors are more (less) conservative than users required when the relevant evidence was precise (ambiguous). Similarly, Salterio and Koonce (1997) provide evidence that auditors' treatment of clients' capitalization versus deferral decisions depends on whether the relevant precedents unanimously favor one alternative. When the precedents favor one alternative, auditors follow the precedents, but when the precedents are mixed, auditors tend to follow their client's preference. Mayhew, Schatzberg, and Sevcik (2000) provide consistent evidence in experimental markets. When participants in the role of auditor were sure of the appropriate disclosure, they made that disclosure, but as their uncertainty about appropriate disclosure increased, they tended to misreport in favor of their client.

Other studies have focused on the role of specific incentives in auditors' reporting decisions. For example, Hackenbrack and Nelson (1996) provide evidence that auditors are more

likely to allow their clients to take aggressive accounting positions if the auditors' litigation risk is reduced, and that auditors justify the aggressive position with aggressive interpretations of the relevant financial accounting regulations. Hackenbrack and Nelson hold constant the underlying audit evidence while varying auditors' incentives and whether those incentives favored accrual or footnote disclosure of a contingency, allowing them to infer with high confidence that incentives were driving the effects they observed. Using the same case materials, Kennedy, Kleinmuntz, and Peecher (1997) provide evidence that, even when litigation risk is relatively high, auditors may tend to take aggressive reporting positions when they can diffuse personal responsibility by consulting other experts within the firm. Wilks (2001) provides evidence that auditors' interpretation of evidence and decisions are affected by the views of more senior auditors. Beeler and Hunton (2001) provide evidence that incentives from lowballing or management-advisory services affect audit partners' going concern judgments. Bazerman, Morgan, and Loewenstein (1997) suggest that auditors cannot be independent because of the unconscious effect of such incentives, or even because of a sense of auditor-client affiliation that occurs through multiple interactions. However, Dopuch and King (1996) provide evidence that competitive pressures can reduce the effect of incentives like lowballing, and King (2001) provides evidence that, holding constant economic incentives, professional-group affiliation can offset the influence of auditor-client affiliation, demonstrating that offsetting affiliations can have offsetting effects on auditors' independence.

A smaller group of studies examines how *managers'* incentives affect the aggressiveness of their reporting decisions. These studies take two approaches. One approach is to elicit managers' judgments directly. For example, Cloyd, Pratt, and Stock (1996) gather data from corporate financial executives at both public and private manufacturing firms. They provide evidence that, when a manager has selected an aggressive tax treatment, the manager tends to choose a financial accounting method that conforms to the tax choice in hopes of better defending the appropriateness of the tax choice if it is later questioned by the IRS. Managers of public firms were less likely to choose conformity than were managers of private firms,

presumably because managers of public firms face more disincentives for making income-decreasing financial accounting disclosures.

The second approach is to elicit the joint product of the manager-auditor negotiation indirectly from auditors. Three different studies use different versions of this approach. Libby and Kinney (2000) manipulate factors that affect managers' incentives and ask auditors to determine how the audited financial statements would appear. They provide evidence that correction of quantitatively immaterial errors is much less likely if the correction would cause the firm to miss analysts' EPS forecasts (i.e., is qualitatively material), and that the recently promulgated SAS 89 has little effect on this behavior. Gibbins, Salterio, and Webb (2000) develop a model of auditor-client negotiation and support their model by surveying auditors concerning their experiences negotiating contentious accounting issues with their clients. Nelson, Elliott, and Tarpley (2000) survey auditors concerning their experiences with clients' attempts to manage earnings, and provide evidence concerning managers' incentives for attempting earnings management, the financial accounting areas in which managers attempt earnings management, and the circumstances under which auditors pass or thwart managers' attempts.

Overall, these studies provide direct evidence that managers and auditors use the flexibility inherent in accounting rules to make disclosures that are favored by their incentives. Holding constant amount of flexibility, changes in incentives move disclosure in the direction favored by those incentives. Holding incentives constant, increasing flexibility increases the degree to which incentives affect decisions.

Certainly one direction for future research is to continue examining how managers' and auditors' incentives affect their decisions. In addition, the literature could work more to identify the processes through which these effects occur. To what extent are these effects intentional and strategic versus the unintended results of cognitive limitations? Wilks (2001) provides evidence that incentives affect decisions more when the incentives are made apparent to subjects prior to evaluating evidence, suggesting that incentive effects influence the evaluation process as well as

the decisions that result from that process. Beeler and Hunton (2001) provide evidence that incentives affect both the favorability and weighting of evidence, and that auditors believe that incentives affect other auditors' judgments, but not their own. A fruitful direction for future research is to further understand how and when such incentive effects occur. Another useful direction is to examine how changes in regulations or other interventions might affect the aggressiveness of financial reporting. For example, Libby and Kinney (2000), Hirst and Hopkins (1998), and Maines and McDaniel (2000) provide evidence of recent regulatory changes that do not appear to prevent managers from making aggressive reporting decisions. Cuccia, Hackenbrack, and Nelson (1995) provide evidence in a tax context that increasing the precision of a standard does not prevent aggressive reporting when the underlying evidence also provides latitude for interpretation. When coupled with evidence of the effect of incentives on reporting judgments, findings indicating the ineffectiveness of some regulatory interventions suggest that regulators might reduce aggressiveness more effectively by addressing incentives directly via changes in penalties. Alternatively, other approaches like improvements in audit-evidence sequencing (Phillips 1999) or within-firm consultation (Kennedy et al. 1998) might also affect the aggressiveness of financial reports, by affecting the extent to which auditors discourage aggressive reporting.

Finally, future research could focus more on the *interaction* among participants in the financial reporting process. Researchers are only beginning to consider the process by which auditors negotiate with their clients to produce the joint product that investors consume. Also, the increasing role of audit committees in this process remains largely uninvestigated. Addressing these issues via experiments (e.g., Libby and Kinney 2000), surveys (e.g., Gibbins et al. 2000; Nelson et al. 2000), and laboratory markets (e.g., Mayhew et al. 2000) appear to be useful directions for future research. These issues are discussed more in section 3.4.

3.2. How do information users interpret reports, given their knowledge of the regulations governing those reports, and their knowledge of the reporters' incentives?

Three streams of literature address distinct facets of this question:

(1) How do accounting methods and disclosure alternatives affect earnings predictions and value estimates of investors and information intermediaries?

(2) How do investors and analysts use the time-series properties of earnings to predict future earnings?

(3) What determines analysts' forecasting and valuation performance?

We discuss each in turn.

3.2.1. How do accounting methods and disclosure alternatives affect earnings predictions and value estimates of investors and information intermediaries?

The earliest experimental research in financial accounting tended to be motivated by the need for evidence to address specific accounting policy debates. These studies focused on *whether* investors and others adjusted appropriately for the effects of accounting methods and disclosure alternatives (e.g., Dyckman 1964; Jensen 1966). Looking back on the earlier literature, it is readily apparent that the answer to this question is "*sometimes*." Some participants in nearly every study of this type demonstrate some degree of functional fixation; they do not fully adjust for differences in the effects of accounting alternatives on the bottom line (see Maines 1995, p. 90, 91). As a consequence, firms that are in identical economic circumstances except for their choice of accounting alternatives are sometimes judged to be different.

These specific policy-oriented studies did little to tell us how the extent of functional fixation will vary across types of decision makers or economic circumstances, or what psychological processes underlie insufficient adjustments to accounting policies. Consistent with this concern, much recent research has heeded the advice of Maines (1994) to focus on the dimensions of disclosure, environmental factors, and processes that determine the degree to which appropriate adjustments are made. In response to a recent call for more specific policy-oriented experiments (Beresford 1994), Maines (1994) noted that "Psychological and sociological research may be most productively used to guide behavioral accounting research on general issues that underlie many different accounting standards, rather than focusing on issues

relevant to only one standard." Understanding the effects of these general factors will dramatically broaden the relevance of this research.

Three groups of studies demonstrate progressive refinement in the manner in which this research question has been addressed. The first group focuses on the mechanisms through which placement and classification of accounting disclosures affect the use and interpretation of the disclosures. The second group explicitly or implicitly recognizes that managers issuing accounting reports have their own strategic interests and will report opportunistically, and examines how users respond to voluntary disclosures by managers. The third recognizes that analysts' respond to their own strategic interests and examines how users respond to potential relationship induced bias in analysts' reports. We discuss each in turn.

General Issues Underlying Functional Fixation. The development of category structures in memory plays a major role in allowing expert decision makers to respond effectively and efficiently in complex decision environments. In these structures, attributes are associated with categories as opposed to individual instances of the category. An individual instance or event is then interpreted based in part on its category membership. This allows for efficient and often effective processing of attributes of the environment, but sometimes produces errors when the particular instance does not match the typical category attributes well. A number of recent papers have recognized that classification issues like the assignment of a financial disclosure to a particular financial statement, to a specific subsection within a statement, or to the notes, will affect decision makers' categorization of that disclosure and interpretation of its relevance and meaning.

Existing studies have examined three dimensions of classification. Hopkins (1996) examined the effects of classification of items on the right side of the balance sheet as debt, equity, or mezzanine financing on judgments of the stock price effects of new financing. He found that experienced buy-side analysts who had knowledge of the differential stock price effect of debt and equity issuances found in financial economics research responded to the issuance of hybrid securities based on their categorization. When the securities were classified

as mezzanine, for which the analysts had no well-defined category, they responded based on the attributes of the individual security. Similarly, Hopkins, Houston, and Peters (2000) examined issues related to categorization of costs as operating expenses, one-time charges, or note disclosure. Experienced buy-side analysts treated the accounting acquisition premium in a merger in part based on its classification. One-time charges and note disclosures were treated as less relevant to stock valuation than operating expenses. Finally, Hirst and Hopkins (1998) and Maines and McDaniel (2000) examined whether placement of elements of comprehensive income on the income statement versus the statement of stockholders' equity affected the ability to detect earnings management and changes in earnings volatility. Information placed on the income statement (the primary performance statement) was much more likely to be treated as relevant to future performance estimates by the experienced analysts in Hirst and Hopkins (1998) as well as by the evening MBA students in Maines and McDaniel (2000).

Maines and McDaniel (2000) also present the beginnings of a theory of format effects. Their theory lists five factors that affect the degree to which investors will rely on a particular disclosure in assessments of corporate performance: placement, labeling as income, linkage (to net income), isolation, and degree of aggregation. Such a theory holds the promise of allowing predictions of effects beyond the scope of individual studies, as Maines (1994) recommends. Future research can refine and test the model in other circumstances.

Other studies identify the stage in the decision process where any failure to adjust for accounting or disclosure differences occurs. Following prior credit analysis and auditing research (e.g., Abdel-Khalik and El-Sheshi 1980; Bonner 1990), Lipe (1998) uses a series of debriefing questions to separate the effects of measurement from weighting. She examines whether investors can accurately assess the variance and covariance of returns in making risk assessments and whether they use those assessments in their investment decisions.[7] Maines and McDaniel

[7] She also examines how they react when market and accounting measures conflict. Her study is unique at this point in jointly examining the role of accounting and non-accounting information. It also suggests the possibility that the weight placed on normatively relevant information may change with the inclusion of less-relevant information and presents a potential explanation for the lack of diversification of individual portfolios.

(2000) use a combination of debriefing questions and regression analysis to determine whether differences in accessing the information cues, interpreting or measuring the cues, or weighting the cues caused their results. They suggest that participants in all disclosure conditions accessed and interpreted the cues in the same manner, but weighted them more heavily in the income statement presentation condition.

Another set of studies uses improved theories of functional fixation to define "superior" disclosure methods. Early studies only determined if different judgments or decisions are made and ignore the issue of determining the superior disclosure method. Many of the newer studies specify subtasks necessary for successful final judgments or decisions, such as detection of earnings management (Hirst and Hopkins 1998), assessment of variability in underlying "core" earnings (Maines and McDaniel 2000), or covariance assessment (Lipe 1998). Alternatively, Maines et al. (2000) approach the question of assessing which disclosure method is superior in a way similar to the training and decision aids literature in auditing. They suggest that high quality reporting methods (1) allow novice decision makers to perform like expert decision makers and (2) allow the same decisions to be made as completely disaggregated disclosures. They apply their approach in a study of joint-venture financial reporting standards. The approach is consistent with the SEC and FASB's concern for the naïve investor, as well as efficiency concerns and Hand's (1990) suggestion of investor sophistication effects as a partial explanation for market inefficiencies. This study, Maines, McDaniel, and Harris's (1997) study of segment standards, and a number of the above-mentioned are motivated in part by a particular policy issue of current interest. Again, we believe that their impact is determined by their ability to relate the particular policy issue of interest to more general phenomena that inform a wider array of policy questions.

Responses to voluntary disclosures. The studies discussed above implicitly assume that disclosures are generated by a neutral process. However, managers issuing accounting reports generally have their own strategic interests and will report opportunistically. A number of studies address how this strategic element affects users' decisions.

The first two studies examine the effects of the form of disclosures. Kennedy, Mitchell, and Sefcik (1998) examine how investors interpret the different allowable forms of contingent environmental liability disclosure: minimum, best estimate, maximum, or range of the distribution. Experienced financial executive, manager, banker, and MBA student participants' assessments of the distribution of possible losses implied by each disclosure did not match the commonly accepted meaning of the terms. For example, when the "best estimate" was disclosed by management, the participants interpreted it as the minimum, and when a range was disclosed, the participants' estimate of the expected value was well above the midpoint of the range. The participants clearly believed that managers bias their disclosures downward.[8] It also indicates that accounting information has different effects on different judgments, in this case, management credibility and firm value.

Hirst, Koonce, and Miller (1999) examine investors' interpretation of point versus range forecasts and historic forecast accuracy on earnings forecasts and confidence in forecasts (which they relate to trading). If both of these forecast attributes indicate precision of the forecast, they both should affect forecasts and confidence. However, only prior accuracy had an effect on earnings forecasts, while both factors affected confidence and trading. This again indicates that normatively relevant attributes of accounting information may affect some judgments and decisions but not others.

Libby and Tan (1999) and Tan, Libby, and Hunton (2000) investigate the effects of earnings warnings or preannouncements on sell-side analysts' forecasts of future periods' earnings. Libby and Tan provide a demonstration of the process through which the same disclosure can have differential effects on different judgments and decisions. They examine why analysts say in the press that they reward firms that warn, yet punish them in their forecasts. They demonstrate that this inconsistency results from the simultaneous processing of the warning

[8]Participants also believed that managers that decided to disclose the minimum were the least credible, yet they valued their firms the most highly. This suggests that the accounting standard provides managers with a perverse incentive to provide the least informative disclosure.

and earnings announcement in answers to press questions versus the sequential processing of the same signals in the forecasting setting. Tan, Libby, and Hunton (2000) demonstrate that firms that low-ball preannouncements of both positive and negative earnings surprises will receive higher forecasts for future period's earnings, even though the reporting managers themselves are judged as having lower integrity and competence. Also, analysts are aware of management's tendency to low-ball the preannouncements, but do not adjust their estimates of earnings of first time preannouncers in light of this base rate knowledge. This again indicates that known attributes of accounting information do not affect all judgments in the same manner.

Responses to analysts' forecasts. Hirst, Koonce, and Simko (1995) and Ackert, Church, and Shehata (1997) investigate the effects of potential bias in analysts' reports on investors' use of those reports. MBA student subjects in Hirst, Koonce, and Simko (1995) expected analysts whose employers also provide investment banking services to the company to be more biased than those that do not. However, this perceived bias only affected their reliance on the report when the report gave a negative recommendation. Similarly, the strength of the analysts' arguments had an effect only for negative recommendations. Ackert, Church, and Shehata (1997) extend this study to a multiperiod setting where subjects have the option to acquire forecasts from analysts, and also observe actual earnings. Individuals were much less willing to acquire analysts' forecasts that proved to be biased in the past, even when the forecast information was useful. Both studies suggest the need to better understand the processes that determine when reports from analysts and other information intermediaries will be purchased and relied upon.

A general picture emerges from the above studies. First, management's often cited (Beresford 1994) preoccupation with the bottom line, and more specifically with potential penalties for earnings volatility and effects of cosmetic differences, appears at least in part well founded. Second, we have begun to understand that placement, categorization, and labeling all play a role in the simplifications that even professional analysts apply when evaluating accounting information. Future research on the knowledge structures developed by experts for

different types of companies and different types of financial judgments and decisions promises to increase our understanding of these effects.

It is also clear from the above results that the information that decision makers rely upon in their judgments is limited, and the information emphasized clearly changes depending on the financial judgment being made, and other elements of the environment. In fact, awareness of cosmetic differences (and ability to "do the math") does not ensure full consideration of their implications for valuation. The same is true of knowledge of management's tendency to opportunistically employ vague reporting standards or analysts' tendency to bias their reports. There appear to be many cases where the same normatively relevant factors are ignored in one circumstance, but adequately weighted in another by the same decision maker. The fact that results here tie closely to archival data gathered in prior studies adds to the credibility of the results. Future studies should focus on systematically determining the circumstances in which different classes of information receive first-order consideration.

Earlier research on the effect of task complexity on the use of alternative decision rules in credit decisions (e.g., Biggs et al. 1985; Paquette and Kida 1988; see Payne et al. 1992 for a review of psychological studies) will provide some guidance in this area. However, it appears that the determinants of which information items receive first order consideration in particular judgment situations involves more than task complexity. Findings of the importance of cue-response compatibility (Slovic and Lichtenstein 1968) and other task determinants of cue usage in early judgment and decision making research (see e.g., Slovic and Lichtenstein 1971 and Einhorn and Hogarth 1981) may provide useful directions for future research in this area. Furthermore, the interplay between these factors, investor sophistication and effort, and various market attributes discussed in section 3.3 appear critical in determining the importance of cosmetic disclosure differences.

3.2.2. How do investors and analysts use the time-series properties of earnings to predict future earnings?

Post-earnings-announcement drift has become a very active stream of archival research. Bernard and Thomas (1990) provide evidence that drift arises because investors misperceive the time-series of earnings. Specifically, quarterly earnings follow a Brown-Rozeff model, which has two key elements. One element is the autoregressive component—changes from one quarter of one year to the same quarter of the next tend to be positively autocorrelated. The other element is the "moving average" component—the differences between actual and predicted earnings tend to be negatively correlated from one quarter to the same quarter of the next year. Research by Bernard and Thomas (1990) and Ball and Bartov (1996) indicate that investors underestimate both the autoregressive and moving-average components of quarterly earnings; results from Abarbanell and Bernard (1992) indicate that analysts make a similar mistake.

Recent studies have used the advantages of the experimental approach to understand the psychological nature of investors' and analysts' time-series prediction errors. Calegari and Fargher (1997) provides a logical starting point—they attempt to replicate drift in the laboratory, using experimental controls to rule out the possibility that prediction errors are driven by factors other than judgment errors.[9] Just as archival studies focus only on firms with extreme earnings surprises, Calegari and Fargher use time series that exhibit unusually large earnings changes in the most recent quarter. Their results are largely consistent with archival research—both individual traders and market prices underreact to earnings surprises.

Maines and Hand (1996) extend this finding in two ways. First, they present MBA students with two different 40-quarter time-series. One series has strong autoregressive and moving-average components. Another is simply a seasonal random walk with no such components. Subjects under-react to both elements when they are present, but also act as if the autoregressive element is present when it is not. This suggests that drift may arise in the target

[9] For example, investors and analysts could appear to make prediction errors in archival studies because they respond to information other than earnings, because they have incentives for something other than prediction accuracy, or because they are attempting to manage risk.

environment simply because it is too difficult for investors to discern the autoregressive and moving average terms. Drift may therefore be less severe for firms that adhere more closely to a seasonal random walk. Second, Maines and Hand directly test Bernard's (1993) hypothesis that investors anchor too strongly on earnings from the same quarter of the previous year, perhaps because it is stressed in the reporting format used in the popular press. Maines and Hand test this supposition by presenting a new set of subjects with a Brown-Rozeff time-series, and reporting earnings relative to earnings from four quarters ago. The results raise doubts about Bernard and Thomas's (1990) hypothesis, because these subjects place even more weight on the autoregressive component of the time series. These results suggest the need to test for alternative causes.

Bloomfield, Libby, and Nelson (2000a) argue that drift may arise because people naturally over-rely on unreliable information (Griffin and Tversky 1992; Bloomfield, Libby, and Nelson 2000b), and old earnings numbers tend to be unreliable predictors of future earnings, once more current earnings are known. They test this hypothesis by manipulating information about old earnings performance, holding recent earnings performance constant. Student subjects rely much too heavily on old earnings numbers, and generate errors consistent with post-earnings-announcement drift, even when they are presented with a time series that is much simpler than that used in other experiments. This suggests that drift may not arise merely because the time-series properties of earnings are so complex.

Future research in time-series perceptions might follow several directions. One direction is to integrate the different research approaches described above. The realistic time-series used by Calegari and Fargher (1997) and Maines and Hand (1996) allow them to generalize their results readily to archival settings, but make it difficult for them to ascertain how aspects of the time-series data interact with psychological processes to cause prediction errors. The simpler time-series data used in Bloomfield, Libby, and Nelson (2000a) poses precisely the opposite problems. Future research might attempt to work toward the middle of these two approaches,

either by using time-series that are progressively simpler than in the former studies, or progressively more complex than in the latter study.

Future research might also investigate the model of Barberis, Shleifer and Vishny (1998). That model assumes that earnings follow a random walk, but that investors believe that earnings switch between regimes of positive autocorrelation and regimes of negative correlation. This misperception results in both underreactions to recent earnings changes and overreactions to long-term trends. While such misperceptions are broadly consistent with psychological findings indicating representativeness and conservatism biases, no single study supports its assumptions, and their predictions are not entirely consistent with archival evidence (e.g., Lee and Swaminathan 2000).

Finally, future studies might attempt to integrate research on time-series predictions with other research streams that consider earnings prediction more broadly. For example, how might knowledge of earnings components (accruals, cash flows) alter subjects' time-series predictions?

3.2.3. What personal and process attributes determine analysts' forecasting and valuation performance?

As Maines (1995) notes, a number of studies in the 1970s and 1980s examined the manner in which expert and novice analysts process accounting information (e.g. Pankoff and Virgil 1970; Slovic, Fleissner, and Bauman 1972; Wright 1977; Mear and Firth 1987). The studies assessed various characteristics of information search, cue weighting, judgment consistency and consensus, and self-insight into information processing. A number of the more recent studies in this group used detailed process tracing techniques in an attempt to tie individual or process attributes to judgment accuracy (e.g., Bouwman 1984; Biggs 1984; Anderson 1988). However, most studies were only able to relate process attributes to experience because of subject sample constraints or difficulty in measuring judgment performance. These earlier experiments also did not focus on the effects of analysts' incentives, which have received a great deal of attention in recent archival studies.

Three recent studies have added substantially to our understanding of the relationship of personal and process variables to forecast accuracy as well as the impact of relationship incentives on bias in forecasts. Hunton and McEwen (1997) emphasize both process measurement and disentangling variables that are confounded in natural settings. They address whether sell-side analysts' search strategy and incentives (in the form of their relationship to the company) affected the accuracy and bias of their earnings forecasts. Information search strategy was assessed with an eye movement measurement system that eliminates most concerns about the reactivity and validity of verbal protocols. The authors measured the accuracy of the forecasts made in the experiment as well as historical accuracy from company archives, which assures external validity. Analysts that followed a more directed (as opposed to sequential) search strategy were more accurate both in the experimental task and in practice. The analysts in the underwriting condition gave higher (more biased) forecasts than those in the following condition, which were higher than those in the no relationship condition. Careful use of controls eliminates concerns about omitted variables such as information availability, time on task, and some forms of selection, that could have explained similar findings in archival studies (see Kothari 2000 for a review).

Few studies have examined the knowledge and abilities that lead to successful performance by analysts. Ghosh and Whitecotton (1997) present evidence that two standard psychometric measures of information processing ability (perceptual ability and tolerance for ambiguity) were correlated with forecast accuracy. But, as in Hunton and McEwen (1997), experience was unrelated to accuracy. However, Whitecotton (1996) reports that experienced analysts outperformed MBA students, who outperformed undergraduate students, though the experienced analysts were the most overoptimistic.

Like similar work in auditing, these findings are potentially relevant to the selection and training of analysts, as well as the interpretation of their forecasts and reports. Again, the fact that results here tie closely to archival data, gathered either in the same study in the case of Hunton and McEwen's (1997) accuracy measures, or in prior studies in the case of their

incentives findings, adds to the credibility of the results. Recent archival studies by Mikhail, Walther, and Willis (1997), Clement (1999), and Jacob, Lys, and Neale (1999) have documented differences in the experiences of more and less accurate analysts that may indicate directions for future research. In the auditing literature, expertise studies have refined such findings in studies that specify the knowledge necessary to complete various tasks, when it is acquired, and the mechanisms through which knowledge content and structure affect performance. These studies can provide guidance for future financial accounting research in this area. Other recent work has begun to look at how these individual responses affect market-level performance and the characteristics of markets that will affect information dissemination. This research is discussed in the next section.

3.3. How do individual responses to information affect market-level phenomena?

Early experimental research in financial accounting implicitly assumed that individual behavior would affect market-level prices in some straightforward manner (e.g., the price might be simply the average of all investors' beliefs), and that some investors would lose money to more sophisticated investors by trading unwisely at market prices. Counter-arguments by proponents of the efficient markets hypothesis have led many experimental researchers to make these assumptions explicit and subject them to testing. We divide this literature into three lines: those that address differences between individual and aggregate behavior, information aggregation, and excess trading volume.

3.3.1. Differences between individual and aggregate behavior.

A number of papers examine whether or not individual responses to information extend to the market level. Two papers examine whether individual responses to risk extend to the market level. Coller (1996) shows that both individual traders and market prices respond to uncertainty in public disclosures in a manner roughly consistent with Bayesian rationality. Bloomfield and Wilks (2000) show that, consistent with theoretical and archival work on disclosure, more accurate disclosures increase individual and market prices relative to expected

values, and also increase individual and market liquidity. A larger number of papers show that *biases* in individual decisions result in biased market prices as well. For example, Calegari and Fargher (1997) show that post-earnings-announcement drift persists in a double auction market, and Bloomfield, Libby, and Nelson (2000a) show that overreliance on previous years' earnings persists in a clearinghouse market. Tuttle, Coller, and Burton (1997) show that recency effects extend to the market level.

Dietrich, Kachelmeier, Kleinmuntz, and Linsmeier (2000) conduct a study closely related to the functional fixation (e.g., Hopkins 1996) and voluntary disclosure (e.g., Kennedy et al. 1998) studies discussed in section 3.2.1. They demonstrate that more explicit disclosure of accounting information about oil-producing properties leads to more efficient market prices even though the same information can be inferred from the balance sheet and income statement. Different disclosure forms either mitigate or exacerbate biases in prices. The authors test their process explanation by tying individual participant's behavior to prices to ensure that the market price results are the result of individual information processing biases.

Other research investigates *how* competitive forces might allow less biased traders to have more influence on price, and use that explanation to guide examination of *when* this is more likely to occur. Of particular interest is the "smart-trader" hypothesis, which states that traders who are less susceptible to the bias trade more actively than other traders, driving prices to unbiased levels (Camerer 1987, 1992). The intuition behind this hypothesis underlies the strong-form of the efficient markets hypothesis, which states that prices will fully reflect information even if it is held only by a small number of traders.

Anderson and Sunder (1995) provide evidence that the smart-trader hypothesis might be more predictive among professional traders than among student traders. They compare the extent of base-rate neglect in markets involving student subjects with the bias in markets involving professional traders. They report that price biases in markets of professional traders exhibit less base-rate neglect over time, while price biases in markets of students do not. This is so even though the professional traders' individual value estimates do not appear to differ from

the students' estimates. This suggests that the professional traders are able to trade in a way that reduces bias more (or increases it less).

Bloomfield, Libby, and Nelson (1996) provide evidence favoring the smart-trader hypothesis in a market in which security values are determined by the answer to general business knowledge questions. Traders with more accurate answers do indeed trade more actively than other traders. When prices are influenced by trading volume, prices become more accurate than the simple average of all traders' value estimates. (Prices are no more accurate than average estimates when they are not influenced by trading volume). This study might support the smart-trader hypothesis more strongly than the studies above because inaccurate traders are not biased, but merely uninformed. It is possible that uninformed people are more likely to know that their answers are inaccurate (and therefore trade less aggressively) than biased people, because biases are unconscious.

Kachelmeier (1996a) uses an analysis of bids and asks to show the difficulty in determining exactly how markets can debias prices. He induces a sunk-cost fallacy that significantly increases sellers' asking prices and buyers' bidding prices. However, these biases have no effect on transaction prices, because the higher bids and asks cause more trades to take place at the bids, which keeps prices low.

Other recent studies show that market structure can be important in determining when the smart-trader hypothesis is likely to be supported. Ganguly, Kagel, and Moser (1994) present student subjects with a problem that leads to base-rate neglect. They find that, because traders are not allowed to sell shares they do not own (short-selling is prohibited), market prices are set by the traders with the highest valuation. As a result, market prices exhibit base-rate neglect most strongly (weakly) when the biased prices are higher (lower) than the Bayesian expected values. Bloomfield and Wilks (2000) find strong individual evidence of an "endowment" effect—inconsistent with Bayesian optimization, traders choose higher ask (selling) prices for riskier securities, even as they simultaneously enter lower bid (buying) prices. However, higher risk does not cause the *market* ask price to rise. This form of irrationality at the individual level is

eliminated at the market level because the market ask is determined by the lowest individual ask. The market ask therefore reflects the selling price of the investor who succumbs least to the endowment effect. In this way, the structure of the market combines with the nature of the bias to mitigate the bias at the market level.

Future research could examine the foundations of the smart-trader hypothesis more directly. In particular, what factors might induce less-biased traders to exploit biases, or keep them from doing so? What factors might make more-biased traders curtail their trading activity? How might changes in market structure, or the degree of market depth and liquidity, affect bias mitigation? (Archival studies routinely show larger biases in less liquid stocks.) Future research could also examine how the nature of financial accounting information will affect the difference between individual and aggregate behavior. To the extent that information induces biases, rather than degrees of informedness that differ across traders, prices would seem more likely to represent an average of all traders' beliefs.

3.3.2. Information aggregation and underreaction.

A different stream of research examines the ability of financial markets to aggregate information held by different traders. Like studies of the smart-trader hypothesis, aggregation studies are motivated by the belief that traders who know a security value does not reflect their own information will trade aggressively to exploit that fact, thereby revealing their information to the market.

Early studies on information aggregation showed that markets do often aggregate information. They do so most effectively when security values are tied to states of nature in very simple ways (Plott and Sunder 1988; O'Brien and Srivastava 1991), and when experienced traders have common knowledge regarding the information environment (Forsythe and Lundholm 1990).

More recent studies have examined how uncertainty affects information aggregation. In a series of double-auction markets, Lundholm (1991) manipulates the "aggregate uncertainty"

that remains after combining investors' information about security value. He finds that markets with aggregate uncertainty aggregate information much less efficiently than those with aggregate certainty. Imperfect aggregation can lead markets to underreact to information, because prices will be too high when the aggregate information indicates a very low value, and too low when the aggregate information indicates a very high value. Bloomfield (1996a, 1996b) shows a similar type of underreaction in a setting which allows aggregate certainty, but in which the information structure is sufficiently complex that information aggregation is still very difficult.

Other papers show that market prices can even underreact to public information that need not be aggregated. Gillette et al. (1999) construct a market in which security values are determined by a sequence of random dividends. The authors analyze the market's reactions as the dividends are announced publicly one-by-one. They find that the individual traders' estimates of value underreact slightly to the dividend announcements, possibly because they erroneously believe that random events tend to reverse over time (the "gambler's fallacy"). More interesting is the fact that market prices underreact substantially more than individual value estimates. The reason for this sluggishness in market prices is not clear, but the authors replicate it in both double-auctions and call markets. Bloomfield, Libby, and Nelson (2000b) also observe a similar effect in clearinghouse markets. Bloomfield (1996a) shows that markets react to a public signal when it is subject to manipulation by a self-interested seller, but not when the signal is purely random. These results raise the possibility that post-earnings-announcement drift and underreactions to other information (e.g., fundamental values, analysts' estimates) may arise simply due to a generic underreaction of market prices to information, rather than information-specific biases.

Several future directions for research in this area entail making endogenous the distribution of information among subjects. All of the aggregation studies described above manipulate information distribution by exogenously altering who is given information and who is not. Future studies might relax this assumption by recognizing that collection of information is an intentional action that is driven in part by the perceived benefit of becoming informed, as in

Tucker (1997). Alternatively, one might recognize that some information may be effectively widely distributed because it is more easily analyzed. For example, Sloan's (1996) archival evidence that prices are too high (low) when firms have high (low) accruals might simply reflect an underreaction to financial statement information that is not widely known. This result is consistent with Bloomfield and Libby's (1996) finding that laboratory markets respond more strongly to information that is more widely available. However, a more direct test of this hypothesis would be to give all traders the same information (e.g., a complete financial statement), and vary the ease with which the information can be analyzed (as in Dietrich et al. 2000), as well as the traders' knowledge and training that would help with such analysis.

More generally, researchers might start with the features we argue are essential for progress in functional fixation research—explicitly understanding how people process and interpret the information in financial statements, and then considering how differences in that processing might alter market behavior.

3.3.3. Trading Volume.

A third line of research examines the determinants of trading volume in laboratory markets. Many of these studies are motivated by a generalization of the "no-trade" theorem (Milgrom and Stokey 1982), which shows that under fairly general conditions, information releases should not induce any trading between traders. The intuition is that if one trader expects to make money trading at a given price, the trader on the other side of the transaction must expect to lose money (since trading is a zero-sum game).

Gillette et al (1999) find routine violations of the no-trade theorem: trading volume is generally quite high, and is even higher after very high or low dividend announcements. These results are consistent with archival evidence on trading volume (e.g., Bamber 1987; Bamber, Barron, and Stober 1997), which have generated a number of theoretical models that generate trade through complex interactions between public and private information (e.g., Kim and

Verrecchia 1994). However, the simplicity of the market in Gillette et al. makes such explanations unlikely.

Excess trading is a puzzle in Gillette et al., but it has few welfare implications because all traders are identical, and therefore wealth transfers can be ignored (or are at best impossible to interpret). Bloomfield, Libby, and Nelson (1999) examine excess trading that has very clear welfare implications. They create markets in which less-informed traders hold a subset of the information available to better-informed traders. Less-informed traders unwisely trade with—and lose money to—the more-informed traders. However, additional instructions that clarify to less-informed investors the extent of their informational disadvantages reduce these wealth transfers (although it has no apparent effect on price biases). These results have regulatory implications: less sophisticated individual investors (who have less information than more sophisticated individuals or institutional investors) can be protected by regulations that emphasize the extent of their informational disadvantage.

There appear to be a number of open questions related to trading volume. Archival papers have examined volume in response to earnings announcements, or tie volume to pricing anomalies (Lee and Swaminathan 2000; Swaminathan and Lee 2000). These findings may be caused by factors indicated in economic models (e.g., Kim and Verrecchia 1994) or by psychological factors. The literature on motivated reasoning seems particularly promising, because it examines how initial variations in beliefs and preferences can be magnified by ambiguous public disclosures of information (see Wilks 2001).

3.4. How do strategic interactions between reporters and users of information affect reporting and market outcomes?

Game theory has been exceptionally useful in modeling the strategic interactions between sellers (who can make reports about their value) and buyers who rely on those reports in making their trading decisions. These models potentially have regulatory implications, because they show that seemingly reasonable regulations may be unnecessary or unwise when one considers the joint response of buyers and sellers to the regulation. The models are very difficult to test

with archival methods, because their predictions are derived in settings that are far simpler than natural markets. However, a number of experimental researchers have chosen to examine behavior in settings that closely resemble those described in the models. In this section, we briefly review some of these experiments.

One line of research examines voluntary disclosure models, in which sellers choose between honestly disclosing the exact value of the security they are selling, and not disclosing anything at all. Two papers by King and Wallin find strong support for the qualitative predictions of the models of Jung and Kwon (1988), and Wagenhofer (1990). King and Wallin (1991b) find that increasing the probability that the seller is informed leads sellers to disclose more often, and also leads buyers to draw more unfavorable inferences when they do not observe disclosure (making disclosure a wise strategy for sellers). King and Wallin (1995) show that disclosure is also limited by introducing a cost to disclosing favorable information (a competitor who will take advantage of favorable disclosures to enter the sellers' product market), because even high-value firms might choose not to disclose. In both cases, however, results deviate substantially from the point predictions of the models.

Forsythe, Lundholm, and Reitz (1999) show how disclosure regulations affect the welfare of buyers and sellers in a simple market with voluntary disclosure. When sellers are not permitted to disclose their information about value, many surplus-enhancing transactions do not occur, and both buyers and sellers suffer. Allowing sellers to disclose any value (even a false one) increases market surplus, but these gains accrue almost entirely to the sellers. Requiring sellers' reports to include the true value shifts part of this surplus from the sellers to the buyers.

King (1996) examines whether disclosure patterns change when sellers have an opportunity to develop reputations. He permits sellers to report any value they wish, but imposes a cost on buyers when the seller's report is inaccurate. This setting includes two equilibria. In an "inflation" equilibrium, sellers always report the highest value, and buyers pay expected value net of the cost of inaccuracy. In a "reputation" equilibrium, the seller reports honestly, and the buyers believe the reports until the seller reports dishonestly; at that point, the players revert to

the inflation equilibrium. King finds that an exogenous cost for inaccuracy does permit reputation formation, but that the reputation equilibrium arises only in a few cases.

There are several natural directions for research in strategic disclosure. There is certainly no shortage of new disclosure models to test. However, it is probably more important for researchers to begin to delve into how and why various equilibria do and do not have predictive power. Some researchers have begun to do so by asking whether "adaptive" strategies (doing more of strategies that performed better in the past) lead to a given equilibrium. For example, King and Wallin (1995) find little support for an "adaptively unstable" equilibrium that is not the end result of adaptive behavior. Other researchers focus more directly on the players' thought processes. For example, experiments by Bloomfield and Hales (2000) examine how sellers' abilities to form reputations for honest reporting are influenced by buyers' and sellers' expectations of one another's likely behavior and beliefs.

Future research might also begin to integrate disclosure research with the other literatures described in this section. For example, Bloomfield (1996a) integrates the disclosure literature with the information aggregation literature by showing that sellers are willing to pay a fee to inflate a public signal, even though the information available to the market as a whole is unchanged. They are willing to do this because markets tend to react more strongly to information held by more investors.

Researchers might also integrate economics-based disclosure research with the psychology-based literature described in section 3.1. That research focuses on how investors could use financial reporting choices to draw inferences about managers' incentives and information, but ignores the fact that managers should anticipate investors' reactions. On the other hand, the psychology-based research presents a more comprehensive treatment of financial accounting institutions, by allowing managers to choose how to classify and report accounting information. We believe it would be worthwhile—though difficult—to examine fully strategic interactions in more complex accounting institutions. Researchers in financial accounting might

also attempt to integrate game theory and social psychology, as has been done successfully in the auditing context by King (2001).

4. Effective and Efficient Research Design: Methodological Considerations in Experiments

Section 3 presented a number of directions for future experiments. In this section, we discuss how these experiments can be designed to be both efficient and effective. An experiment is *efficient* if it achieves a given level of effectiveness as economically as possible. An experiment is *effective* if it provides evidence of sufficient internal validity that readers should believe the results of hypothesis tests, while being of sufficient external validity that it bears on a significant part of the financial accounting issue of interest.[10] Both internal and external validity are key to effectiveness. An experiment that lacks internal validity fails by providing a misleading indication of the relation between the dependent and independent variable, while an experiment that lacks external validity produces results that are (or at least should be) divorced from the motivation of the study. We do not provide an exhaustive treatment of research design (see Runkel and McGrath 1972, Kinney 1986, and Trotman 1996 for more comprehensive discussions). Rather, we focus on issues that we believe are particularly important or are often misunderstood. Section 4.1 addresses techniques for maximizing effectiveness through careful hypothesis development and research design. Section 4.2 addresses when it is (and is not) possible to improve efficiency by consuming fewer resources without sacrificing effectiveness. We address the number and type of subjects used in the experiment, the payment of monetary incentives, the use of within-subject designs, and the decision to use single-person tasks rather than interactive tasks (such as financial markets or strategic reporting settings).

[10] Internal validity is the degree to which you can be sure that observed effects are the result of the independent variables. External validity is the degree to which results can be generalized beyond the specific tasks, measurement methods, and participants employed in the study.

4.1. Increasing Experimental Effectiveness

We organize our discussion of experimental effectiveness around the predictive validity model (Runkel and McGrath 1972; Libby 1981). This model provides a useful description of the hypothesis testing process, and focuses our attention on the key determinants of the internal and external validity of a research design.

Figure 1 illustrates the predictive validity model as it applies to Hypothesis H1b from Hunton and McEwen (1997) (hereafter, HM). As noted earlier, based on prior theory and evidence HM hypothesized that sell-side analysts' relationship-based incentives would decrease their forecast accuracy. Analysts' relationship-based incentives were operationally defined as a three-level independent variable: an "underwriting relationship" that has a direct impact on fees, a "following relationship" that creates the need for future access to private information, or "no future relationship." HM expect analysts in the underwriting condition to provide the most optimistic forecasts, those who follow the firm to be next most optimistic, and analysts who do not follow the firm to be the least optimistic. They operationally define optimism (the dependent variable) as the analysts' forecast minus the actual earnings outcome. HM also controlled for a number of other potentially influential variables including subject background, experience, time on task, and information availability.

In Figure 1, link 1 depicts the relationship in HM's underlying theory. No theory can be tested directly; rather, a theory is tested by assessing the relationship between the operational definitions of key concepts in the theory (i.e., by assessing link 4). For this test to be valid, the links between the concepts and the operational definitions (links 2 and 3) must be valid, and other factors that might affect the dependent variable (link 5) must be controlled or have no effect. A study's internal and external validity is determined by the validity of these five links. We now discuss ways in which researchers can strengthen each of these links.

4.1.1. Link 1: Theory and Hypotheses

The first determinant of experimental effectiveness is specification of a good research question. A good research question addresses the relation between two or more concepts, can be

stated clearly and unambiguously as a question, implies the possibility of empirical testing, and is important to the researcher and others (Kinney 1986).

Experimental tests of research questions must rely on some theory depicting forces that influence behavior in the experimental setting. Theories may range from highly specific numerical models (such as those derived from economics or artificial-intelligence cognition models) to more general qualitative predictions based on prior evidence (such as systematic evidence that people use a certain heuristic in a given setting). Regardless of its nature, the theory suggests the expected answer to the research question, and serves to guide the many decisions and tradeoffs that must be made during the design and administration of an experiment. Whereas archival researchers analyze data from secondary sources[11], the experimental setting is specifically designed to gather data relevant to the hypotheses. Consequently, all stages of the design of experiments are profoundly affected by the need for a well-formulated research question and hypotheses. In this section, we emphasize four issues that are particularly important in developing good research questions and hypotheses in experimental financial accounting research.

First, the hypotheses must have external validity; that is, readers must believe that the theoretical concepts and the relationships between them capture important aspects of the target environment. Although people often speak of external validity as an aspect of experimental stimuli, we consider it an element of theory as well. If the theory and hypotheses are appropriately capturing relationships among elements of the target environment, an internally valid experiment will test that theory in a manner that generalizes to the target environment. External validity is established empirically by extensions of the research that test additional hypotheses concerning environmental contingencies that define the limits of generality of the initial hypotheses (Trotman 1996).

[11] That is, the data is initially gathered for a different purpose.

For example, HM's research question of "Do sell-side analysts' relationship with the firms they cover decrease their forecast accuracy?" relates an antecedent (analysts' relationships) and consequence (forecast accuracy) that clearly maps into first order concerns indicated by theory and prior evidence. If the experiment operationalizes those concepts well and provides an internally valid test of their relation, it will provide insight into the real-world effect of analysts' incentives on their judgments. Future research can then test the extent to which those insights can be generalized.

Second, experimental research questions in financial accounting should focus on how theories drawn from fundamental disciplines (such as psychology and economics) interact with details of financial accounting institutions (as discussed in section 2.4). As Gibbins and Swieringa (1995) suggest, accounting experiments should be "both theory driven and setting sensitive."

Tying the accounting institution to theory from a fundamental discipline allows hypotheses to have relevance beyond the very specific practice context that motivated the experiment (as recommended by Maines 1994). It also allows experimenters to contribute to both financial accounting and the fundamental discipline. For example, Nelson and Kinney (1997) apply Einhorn and Hogarth's (1986) ambiguity model to predict how ambiguity affects financial statement auditors' and users' judgments of appropriate contingent-liability disclosure. Their study shows how the differences between auditors' and users' incentives lead auditors to use the discretion provided by ambiguous evidence to justify lower levels of disclosure than users desire. This result is of clear interest to financial accounting researchers, and also contributes to psychologists' understanding of how incentives interact with ambiguity.[12]

[12] Of course, the theory should entail some element of doubt before testing. Experiments applying psychology to accounting settings can be uninteresting if readers are certain that the results obtained in psychology will readily extend to accounting even without seeing the experimental results. Experiments applying economics to accounting settings can be uninteresting if they are little more than complex ways of showing that people prefer more money to less (Kachelmeier 1996b).

A more ambitious approach is to use fundamental disciplines to *develop* and experimentally test a general theory that is applied to the financial accounting phenomenon of interest. For example, Maines and McDaniel (2000) identify various general dimensions of formats that signal information importance or that affect the cognitive cost of processing information (see also Lipe 1998). They apply their theory when testing whether information-disclosure format affects consideration of the volatility of unrealized gains and losses, but their theory is much broader than the particular practice context that they examine.

Third, researchers should frame their theories at the *least* specific level that can account for the data expected to arise from the experiment. Stating the theory with greater specificity will simply encourage readers to argue that the results are driven by a slightly different theory (such as a different theory of categorization) that yields identical predictions in the experimental setting. Such debates are rarely productive. If the distinction is likely to be important in accounting settings, researchers interested in accounting issues should consider what other experiments might illustrate this importance. If the distinction is unlikely to have important ramifications for accounting settings, experiments discriminating between such theories are more appropriately seen as contributions to the fundamental disciplines from which the theory is drawn.

Finally, experimental research questions should be based on a theory that describes *causal* relationships between concepts. As discussed above, the key advantage of the experimental method lies in its ability to disentangle factors that are confounded in natural settings, and thus provide indications of how and why phenomena arise. A causal theory also improves external validity, because causal forces are more likely to generalize to different settings. This also leads to a preference for research questions that focus on a *directional prediction of differences*, as opposed to a single point prediction. As Trotman (1996) indicates, "the basis of any experimental design is that one or more independent variables are manipulated and the effect on the dependent variable(s) is observed." Since experiments require abstraction from the real world, any number of differences between the experimental and real-world

environments could affect the particular levels of observed measures. Consequently, evidence consistent with point predictions (e.g., "the market price will be $5.00") and particular parameter estimates (e.g., "managers will weight current year's earnings twice as heavily as prior year's earnings") are unlikely to generalize to real-world environments. Directional effects are more likely to generalize, because differences between the experimental setting and the target setting are more likely to alter the magnitude of an effect than its direction. A focus on directional effects also makes it much easier to design an experiment that controls for competing explanations. We discuss this latter issue further in section 4.1.3.

4.1.2. Links 2 and 3: Operationalizing Dependent and Independent Variables:

Link 2 relates the antecedent theoretical concept A to the independent variable(s) operationalized in the experiment. Link 3 relates the consequential concept B to the dependent variable operationalized in the experiment. An internally valid test requires manipulation of each independent variable in a way that changes only one theoretical antecedent at a time. At the same time, they must construct an operational dependent variable that measures the conceptual variable, and that variable alone. This section discusses three particularly difficult issues in operationalizing variables: (1) choosing the appropriate realism of the stimuli presented to participants, (2) choosing the appropriate levels of independent variables, and (3) using measured independent variables.

Realism of stimuli. A common challenge in operationalizing independent variables is deciding how realistic the stimuli should be. The appropriate level of realism in the operationalization of an independent variable is determined by the role of realism in the theory to be tested.

Experiments testing psychological theories typically present participants with more realistic stimuli than experiments testing economic theory, because psychology-based experiments typically are focused on how participants make decisions using cognitive processes and knowledge that developed in response to their real-world education, training, and

experience. Without relatively realistic stimuli, participants may not rely on the cognitive processes and knowledge of interest. For example, HM's theory relates analysts' knowledge of their incentives to their earnings estimates. In order to test this theory, the experiment must provide the participants with a sufficiently realistic stimulus to activate that knowledge. Similarly, Hopkins (1996) tests the theory that classification of debt-equity hybrid securities alters analysts' inferences about firm value; this theory can be tested only with relatively realistic stimuli and value-assessment tasks.

Experiments testing economic theories typically present participants with less rich information and less realistic stimuli, because they focus on how participants make decisions using economic information given particular preferences, constraints, and incentives. The decision processes depicted in these theories are not hypothesized to depend on task realism, so these studies are less concerned with it. For example, King and Wallin (1991a) test theories relating the probability that a seller knows the asset value to the sellers' disclosure strategies and buyers' responses to those disclosures. That study does not require realism or knowledge of particular real-world institutions, so it uses abstract stimuli and tasks to avoid introducing extraneous factors that might compromise internal validity.

This discussion should not be construed as indicating that all experiments testing theories drawn from psychology (economics) must have high (low) stimulus realism. Experiments testing very general psychological theories (such as the relation between short-term memory and optimism) could contribute to financial accounting research with stimuli and tasks that possess very low degrees of realism. Similarly, experiments testing the effects of superior accounting knowledge on trading profits would require high degrees of realism. It is the goal of the experiment that determines whether realism adds to or detracts from internal validity.

Stimulus realism can also provide benefits beyond that required for an internally valid test of the underlying theory. First, realism can help authors convey to readers the ways in which the results relate to prior research. For example, Hopkins (1996) and Tan, Libby, and Hunton (2000) are able to compare their pricing and earnings-forecast difference results for some

treatments directly to prior archival studies, which increases confidence in the generality of the results of treatment combinations for which no (or insufficient) archival data are available. Second, realism can help subjects understand the task they are being asked to perform, thereby reducing noise in the data. This may be particularly important in economics-based experiments, which place high demands on participants' attention.

However, it is important not to exaggerate the benefits that stimulus realism provides when it is not directly enhancing internal or external validity. Such realism may not substantially increase external validity, which is determined mainly by the theory itself and how effectively the theoretical constructs have been operationalized. Similarly, it is important not to exaggerate its costs. Experimental economists often worry that realism may influence behavior in ways that lie outside their theories, and thus reduce internal validity (Smith 1976, Camerer 1997), but as we will discuss in section 4.1.3, these concerns typically can be dealt with through good experimental design.

Choosing levels of independent variables. After choosing the *nature* of independent variables, the researchers must choose their levels. A general goal is to choose levels that are different enough that the experiment has sufficient power to yield strong effects, yet be within the relevant range.

As indicated above, in some cases it is appropriate to choose levels that depict real-world conditions. For example, HM's independent variable consists of treatment levels that reflect what analysts might experience in practice. Given that their theory is testing the relation between those real-world incentives and analysts' behavior, this realistic depiction provides a strong test of the theory. However, it is usually difficult to ensure a representative sample of independent variable values, which limits the interpretability of levels of effects and parameter estimates in most experiments. Choosing realistic versions of naturally occurring phenomena can also make it difficult to manipulate only a single theoretical antecedent while holding all others constant. This is particularly true in studies of alternative accounting methods or disclosures, where differences in method or disclosure (the experimental treatments in these studies) can convey unintended

information about the nature of the underlying transactions that affect the dependent variable but are not included in the theory being tested. Experimental controls discussed under link 5 can be employed to reduce this concern (see e.g., Hopkins 1996).

In other cases, it can be wise to create levels that are unrealistically extreme. For example, Forsythe, Lundholm, and Reitz (1999) compare a regulatory regime that prohibits disclosure with one that allows any disclosure (even fraudulent statements). While these levels are unrealistic, they allow a very powerful test of effects that would likely generalize to milder changes in disclosure regulations.

It can even be useful to specify at least one level of the independent variables that cannot occur in practice, to enable a cleaner test of the underlying theory. One example of this approach is provided by Libby and Tan (1999). They seek to understand how analysts can say they reward firms for issuing early warning of negative earnings surprises, while actually punishing them in their forecast revisions. Libby and Tan address this question by operationalizing three "warning" conditions. Two conditions are realistic: one in which no warning occurs prior to an earnings announcement, and one in which the warning is followed by the negative earnings announcement. A third condition cannot exist in practice: the warning and negative earnings announcement occur simultaneously. This “simultaneous warning” condition allows them to separate the effect of the warning from the sequential processing of two signals by creating two comparisons (each treatment compared to the simultaneous warning condition) that manipulate only one antecedent. The other two settings enhance external validity by mapping naturally into the institutional setting and archival findings the authors seek to inform.

Regardless of how one chooses the levels of the independent variables, it is usually advisable to conduct manipulation checks. These are measures, often taken during debriefing, which seek to determine whether subjects noticed and interpreted correctly the independent variable(s). Manipulation checks test link 2 of the predictive validity framework. Manipulation checks are particularly useful when analyses reveal no significant treatment effect, since one alternative explanation for the lack of a significant effect is ineffective operationalization of the

independent variable (a link 2 problem). However, it is critical that the manipulation check tests recognition and comprehension of the independent variable, as opposed to serving as another test of the treatment effect. Otherwise, the manipulation check is really just a second measure of the dependent variable (testing link 4 rather than link 2).

Measured independent variables. Some independent variables in accounting experiments are observed, rather than manipulated. Because subjects are not assigned randomly to measured treatment levels, measuring independent variables gives up some of the experimentalist's comparative advantage. Such studies are subject to the same correlated-omitted variables problems that compromise internal validity in archival research. Therefore, it is typically preferable to manipulate important independent variables whenever possible, rather than measuring them.

However, there are at least four circumstances where measuring independent variables is useful. The first is that it is impossible or impractical to manipulate an antecedent. For example, HM hypothesize that analysts that are considered by their firms to be more accurate forecasters tend to use a more directive, hypothesis-driven evidential search strategy. Because HM cannot randomly assign analysts to "high historical accuracy classification" and "low historical accuracy classification" treatments, it is possible that historic accuracy classification is correlated with some other variable (such as age or intelligence) that determines use of a directive search strategy. As a consequence, HM include a number of control variables to test these alternative explanations for results, and are careful to discuss these results in terms of "associations" rather than "causes." A second reason to use measured independent variables is that the theory relating the antecedent to the consequence involves mediating variables (a sequence of links through intervening variables). For example, Hopkins (1996) predicts that the balance-sheet classification of manditorily redeemable preferred stock (concept *a*) affects analysts' beliefs concerning the total amount of equity outstanding (concept *b*), which in turn affects their stock price estimates (concept *c*). Because analysts' beliefs about outstanding equity are actually a dependent variable in a part of his theory (*a* affects *b*), Hopkins cannot manipulate it directly.

Those beliefs become a measured independent variable when testing the second part of the theory (*b* affects *c*).

Similarly, almost every multi-person task involves intervening variables, because the behavior of one person is determined by the (necessarily endogenous) behavior of another. For example, King (1996) tests whether imposing exogenous costs on buyers for inaccurate value estimates induces sellers to report values accurately. One simple breakdown of this theory is that exogenous costs (concept *a*) reduce the prices buyers are willing to pay when the seller has previously reported inaccurately (concept *b*), which leads the seller to choose higher reporting accuracy (concept *c*). Because equilibrium models involve many forces acting simultaneously (e.g., the seller should *anticipate* the buyers' response to his reports, and the buyers should *anticipate* the seller's response to their likely price-setting behavior), it is difficult to measure all of those forces simultaneously in one experiment. Thus, King measured some potential intervening variables (he chose to examine how sellers' reporting accuracy affects buyers' reliance on those reports), but not others.

One way to avoid measured independent variables is to construct separate experiments testing the separate parts of the theory. Hopkins could have tested the "*a,b*" and "*b,c*" links separately or in sequence, reasoning that finding support for both links suggests (but does not demonstrate) an "*a,c*" link. However, he chose to provide a clean test of the "*a,c*" link by testing it directly, and using subsequent measurement of "*b*" to provide comfort that subjects behaved as predicted. Similarly, King could have separately tested buyers' responses to seller decisions. However, we believe that both authors were justified in focusing their cleanest tests on the primary antecedent and consequence concepts in their theory. A full understanding of the causal path may be somewhat encumbered by the problems associated with measured independent variables, but remaining problems can be addressed in future research. For example, Bloomfield and Hales (2000) use a series of experiments to understand more of the linkages in King's study.

Third, it is sometimes much less interesting to examine reactions to a manipulated variable than a naturally occurring one. For example, it would have been less interesting for

Hopkins (1996) to test whether analysts who are *told* that there are more shares outstanding would place a lower value on a firm's stock, all else held equal. It seems much more reasonable to ask whether the same analysts would use that belief to assess stock value when the belief arises naturally. This type of concern is even more salient in tests of equilibrium models.

Fourth, measured variables often provide the keys to understanding underlying processes that produce the effects of interest. For example, Maines and McDaniel (2000) make a contribution by demonstrating effects of format on judgments of management effectiveness and stock risk (an "*a,b*" link), even though their lack of significant effects of format on valuation could be viewed as an insignificant "*a,c*" link. After all, each intervening successive link adds noise and diminishes the experimenter's ability to detect an effect of "*a*" on a later consequence (particularly when the later consequence is a very complicated judgment like stock valuation). Only by eliciting intervening variables does a clear pattern of results emerge. Hirst, Koonce, and Miller (1999) demonstrate the importance of specifying the correct causal path. They show that the form of a forecast will affect trading decisions, not through estimates of future earnings, but through confidence in estimates. This further highlights the need to elicit intervening dependent variables that aid in interpreting results with respect to tests of complex theories. We encourage researchers to measure potential intervening variables whenever possible, if only after they measure their primary dependent variable.[13]

4.1.3. *Links 4 and 5: Statistics and Other Potentially Influential Variables*

As noted earlier, internal validity refers to the degree to which variation in the dependent variable can be attributed to variation in the independent variable. Link 4 assesses the relations between the operational independent and dependent variables. Link 5 captures "other potentially influential" or "extraneous" variables besides the independent variable that could affect the dependent variable. A key advantage of the experimental approach is that the effects of

[13] Of course, the experimenter needs to worry about carryover effects (i.e., earlier measurements affecting later behavior). Sometimes the order in which successive dependent variables are elicited is manipulated between subjects to reduce this concern. This is discussed further in section 4.1.3.

extraneous variables can be controlled for primarily by holding them constant or through randomization. As a result, statistical analyses in experiments are typically straightforward, often consisting of simple t-tests, ANOVAs, or non-parametric equivalents. Extraneous variables can also be measured as in archival studies, and used to enhance the power of analyses by accounting for variation in the dependent variable that is not related to the theory being tested. Finally, extraneous variables can be manipulated to directly test their effect. Given the expense typically associated with this approach, it should only be used when the experimenter believes the extraneous variables cannot be dealt with another way.

Very complex statistics are typically necessary in experiments only when they rely heavily on measured independent variables, or when researchers must try to boost power when subject resources are scarce. When those circumstances are not apparent, complicated statistical tests may signal poor experimental design – the experimenter is trying to grapple after the fact with concerns that should have been headed off with good experimental design.

This section describes some of the powerful array of techniques experimenters can use to deal with extraneous variables. The most important technique available to the experimentalist to control for extraneous variables is to assign subjects randomly to treatments. Random assignment, combined with manipulation of independent variables, enables experimentalists to ensure that their results are not biased by factors of which they are aware, as well as factors of which they are not aware. For example, HM randomly assign analysts to incentive-treatment conditions. This results in an unbiased distribution of industry familiarity, age, experience, prior accuracy, etc. across the three levels of the incentive treatment. Thus, HM can conclude, with a specified level of statistical confidence, that these variables, and other unspecified variables such as motivation or breakfast size, did not account for the results. In fact, had HM not chosen to measure analysts' experience and use it as a covariate to reduce variance in their analysis, they could have ignored experience and expected that it would not affect their mean results because of random assignment across treatment conditions.

More generally, random assignment to treatment conditions allows experimentalists to avoid many of the omitted variable concerns that limit causality inferences in archival studies. For example, Kothari (2000) notes that the direction of cause and effect between relationship and forecast optimism documented in the archival literature is not clear. It could as easily result from managers' selection of investment banks whose analysts provide a more optimistic forecast as from opportunistic forecasting by analysts with relationships. This selection alternative explanation is eliminated in HM by random assignment of analyst subjects. As a result of random assignment, the expected value of analyst optimism *prior* to the treatment is unbiased across the three incentive treatment groups.

Second, experimentalists can hold extraneous variables constant at a particular level. For example, HM hypothesize that analysts who will be underwriting securities exhibit different forecast bias than analysts who do not, because they face different incentives. However, compared to non-underwriting analysts, underwriting analysts could also have larger amounts of information available about a firm, or spend different amounts of time forecasting earnings. HM deal with these potential alternative explanations for changes in their dependent variable (forecast accuracy) by holding constant across treatments the amount of information analysts have available and the amount of time analysts can spend on the experimental task. More generally, experimentalists typically hold constant aspects of the institutional setting that they believe are potentially important but that are not part of the portion of the research question examined in that particular study.

A third way to deal with extraneous variables is to measure them (typically during debriefing). These measurements can be used as covariates or measured independent variables to account for their effects. For example, HM identified prior research that indicated that analysts' forecast accuracy changes as they become more experienced. Since a general experience effect was not part of their hypotheses, but might affect their dependent variable, HM measured experience by eliciting years spent as a financial analyst and used it as a covariate in their analysis. Years of experience cannot have been influenced by HM's treatment effect, so

they use it as a covariate to reduce noise in their analyses without fear that it is actually capturing some element of the effect of the independent variable on the dependent variable (link 4). Similarly, Hirst, Koonce, and Miller (1999) use a pretest measure of forecasted earnings taken before the treatment was administered to reduce noise and increase power.

Measurements of extraneous variables are also useful for testing competing explanations for experimental results. For example, Hopkins (1996) tests whether subjects infer management signaling or differential tax treatment from the balance sheet classification of the hybrid security. Either of these inferences could explain an effect of classification on forecast error, but neither is included in Hopkins' theory. Hopkins provides evidence against these explanations by eliciting in debriefing subjects' inferences about the underlying transaction and demonstrating a lack of significant difference in inference between treatment conditions. Such measures operate much like a manipulation check, but rather than providing evidence that the independent variable operationalizes the antecedent concept the experimenter intended, they provide evidence that the independent variable did not operationalize antecedent concepts other than those intended by the experimenter. The assurance they provide is limited (in that they provide evidence by finding an insignificant difference), but it is assurance nonetheless.

A fourth way to deal with extraneous variables is to manipulate them and test their effects. For example, Bloomfield, Libby, and Nelson (2000b) present their subjects with a number of securities, and vary between subjects the order in which securities are presented. They test for order effects and find none, allowing them to discount order of presentation as a potential explanation for their results. Even if they did not test for such effects, manipulating order in a balanced design would reduce the risk that results are specific to a particular order. In general, manipulating factors unrelated to the hypotheses can be useful, but expensive in terms of use of subjects.

Finally, experimentalists can deal with link-5 factors by ignoring them. By "ignore" we really mean "abstract from," because those factors will not be included in the experimental environment. Ignoring some extraneous variables is necessary because it is not practical to

mimic all elements of reality in an experiment; some abstraction is necessary for the experiment to be conducted in a timely manner. To the extent that subjects make assumptions about information that is not included in the experimental environment, those assumptions are randomly distributed across treatment conditions, and do not affect interpretation of results, as long as the treatments do not differentially affect subjects' assumptions about extraneous variables.

It is important to note that these methods of accounting for extraneous variables are effective *only* when the experimental design manipulates the variables of primary interest to test effects of *directional* predictions. For example, Tuttle, Coller, and Burton (1997) wish to examine how security prices are influenced by the order in which information is revealed to investors. They provide investors with rich firm-specific information about market conditions and corporate events, rather than the abstract information used in many markets experiments. Because the authors cannot know exactly what knowledge investors bring to bear in interpreting this rich information, it could have a number of unknown effects on stock price, and might lead prices on average to be higher and lower than they should be. However, rather than comparing prices to a point prediction of true value, they examine whether the order of information release causes a *difference* in prices. This difference cannot be affected by extraneous variables created by the rich information (although they surely exist), because the total information is held constant across the settings being compared.

As discussed by Bloomfield and Libby (1996), this type of "paired securities" design can generally be used to eliminate concerns about unanticipated effects of realism in experiments. Experiments that attempt to compare behavior to point predictions sacrifice this powerful form of experimental control. Even apparently innocuous variables in an experimental setting (such as the color of a computer screen or the time of day at which data collection occurs) could cause deviations of behavior from a point prediction, but are unlikely to cause those deviations to vary across levels of the manipulated independent variables.

4.2. Increasing Experimental Efficiency (Without Compromising Effectiveness)

Experimenters make many choices that affect the amount of resources consumed by their experiments. This section discusses four such choices: whether to use professional subjects (which are difficult to obtain); whether to provide those subjects with monetary incentives (which are expensive); whether to use between-subjects designs (which use more subjects than within-subjects designs); and whether to place subjects in a laboratory market (which requires more subjects than would a study of individual judgments). Choosing to consume more resources does not necessarily increase experimental effectiveness. Rather, it increases effectiveness in some circumstances, reduces it in others, and has a small enough effect in others that it is not justified from a cost/benefit perspective. We discuss each choice in turn.

4.2.1. Subject Selection

When should experiments use professional subjects? Our advice is to match subjects to the goals of the experiment, but to avoid using more sophisticated subjects than is necessary to achieve those goals.

Experiments that examine the effects of some attribute subjects have developed before entering the experiment must use subjects who possess the necessary attribute. Many studies use experiments to "peer into the minds" of specific groups of experienced professionals to determine what they have learned about relevant concepts and events and how that learning affects decisions. Hopkins (1996) examines how knowledge of the differential effects of debt and equity offerings determines how classification of debt-equity hybrids affects analysts' judgments. Libby and Kinney (2000) seek to explore how auditors' beliefs about managers and their own incentives determine the effect of old and new regulations. In both of these cases, the experimenter is interested in how subjects' use of some type of knowledge learned in the real world causes treatment effects, so they must use subjects with the requisite knowledge. Thus, these studies use professionals as subjects.

In some cases, the experimenter can train student subjects to possess an attribute (e.g., knowledge) that the experimenter is interested in examining. This approach is cost-effective

given students' greater availability than professional subjects, and is well suited for testing the effects of specific features of the learning environment and elements of the resulting knowledge (cf. Bonner and Walker 1994). However, this must be done with care since recently acquired knowledge is unlikely to be of the same depth and breadth, or integrated as well with subjects' pre-existing knowledge.

Student subjects are also entirely appropriate in studies that focus on general cognitive abilities, or responses to economic institutions or financial-market forces that are expected to be learned within the experimental setting. Maines and Hand (1996) provide an example of the former; they examine the effects of general tendencies in the processing of time-series information on forecasting behavior. Any of the reporting studies by King and Wallin (1991a, 1991b, 1995) provide examples of the latter; those studies examine how subjects respond to the strategic forces in disclosure games.

Other experiments focus on the judgments of general or novice investors, and so require subjects who possess only basic familiarity with accounting and investing. Student populations that have such basic familiarity are appropriate here as well. MBA students and executive-program participants are particularly useful, as they often have some accounting knowledge and investing experience. Studies of this type employing student subjects include Bloomfield, Libby, and Nelson (1999, 2000a), Hirst, Koonce, and Miller (1999), Hirst, Koonce, and Simko (1995), Kennedy, Mitchell, and Sefcik (1998), Lipe (1998), Bloomfield and Libby (1996), Maines and McDaniel (2000), and Nelson, Krische, and Bloomfield (2000).

In general, experimenters should avoid using professional subjects unless it is necessary to achieve their research goals. In addition to increasing the experimenters' own time and expense, inappropriate use of professional subjects has negative externalities—they may make it more difficult for other experimenters to gain access to this very valuable resource.

4.2.2. Monetary Incentives

When is it appropriate to provide explicit monetary incentives in financial accounting experiments? As in subject selection, the answer should be driven primarily by the goals of the experiment.

First, as noted above, experiments that focus on incentives rely on participating professionals to bring their knowledge of and behavior learned in response to real world incentives to the experiment. Such experiments attempt to examine how professional practice has provided professionals with incentives that affect their behavior in particular ways. For example, HM studied the effect of analysts' incentives on their forecast accuracy, with those incentives determined by the analysts' perceptions and understanding of the relationship that the analyst has with the firm whose performance is being forecasted. Providing performance-contingent incentives in this type of experiment would distort or interfere with the effects of the real world incentives, and is therefore inappropriate. While the effects of professionals' perceived incentives might be diminished in the experimental setting, their direction should not be altered, so their directional effects should not be altered.

Experiments testing responses to economic theory (such as those described in section 3.4) need to provide performance-contingent incentives in order to induce subjects to possess the incentives assumed by the economic model (Smith 1976). Without such incentives, a fundamental causal element of the model may not be present, and there is no reason to expect theoretical predictions to hold. Performance-contingent incentives are almost always appropriate in laboratory market experiments that examine how individual biases can be mitigated by competitive forces. For example, the "smart trader" hypothesis relies on an assumption that more accurate traders trade more actively because they will earn money by doing so.

A researcher who has concluded that performance-contingent incentives are appropriate must then decide on how sensitive payments should be to variations in performance. Our casual observations suggest that most experimental tests of economic theories pay subjects an average of $8 to $20 per hour, with payments ranging from a $5/hour to $100/hour (or sometimes more).

These numbers reflect tradition and resource limitations more than any reasoned theory. These incentives are obviously much less than most agents in financial accounting target environments would expect. However, we doubt behaviors would be substantially different with larger incentives. Past experiments show little evidence that biases are eliminated by incentive compensation, just as financial rewards have not allowed athletes to run a 3-minute mile. Limitations on abilities, rather than a lack of reward, drive these results. More generally, larger monetary incentives might reduce the size of biases, but are unlikely to alter their basic nature and direction.[14] Thus, larger incentives would probably not change the inferences drawn from directional hypothesis tests.

4.2.3. Within- vs. Between-Subjects Designs.

When should experiments use between-subjects designs, rather than within-subjects designs? Within-subjects (or "repeated-measures" designs) where subjects provide more than one observation, generally enhance statistical power by allowing control of between-subjects differences (i.e., there is a "subject factor" in the analyses that accounts for subject-specific noise). This approach has the added advantage of using fewer subjects. However, repeated measures designs can also affect results by making treatment effects more salient, which may signal to subjects that the experimenter wants them to respond to the manipulation (the familiar "demand effect" concern). Also, repeated measures are vulnerable to carryover effects from the elicitation of one measure to the next. Therefore, these designs are most effective when increased salience of manipulated variables is desirable from the standpoint of the experiment's goals and/or when any carryover effect is desired or can be minimized via manipulation of the order in which measures occur.

As noted earlier, Hirst, Koonce, and Miller (1999) use one type of repeated measures design, the pretest-posttest design. Their subjects first forecast earnings and assess confidence in that forecast, given only company background information and the prior years' financial data.

[14] See Kachelmeier and Shehata (1992) for a study on how very large incentives influence responses to risk.

The subjects were then provided with the experimental treatments (management forecast and information about management forecast accuracy), and again forecasted earnings and assessed confidence. This pretest/posttest design allows Hirst, Koonce, and Miller to increase power by using the pretest as a covariate in their analyses or by analyzing the change in forecasts caused by the treatment. Since they want their subjects to attend carefully to the information contained in their treatments, and their analyses are based on comparisons between treatment conditions (which hold treatment salience constant), they are not concerned about drawing extra attention to the treatment.

Within-subject treatments are particularly common in laboratory markets and games. For example, Bloomfield and Wilks (2000) create a setting in which each group of subjects participates in eight different treatments (every cell of a 2x2x2 design) over the course of two trading sessions. Such repetition reduces noise in the data, which is often high in early repetitions because the environment is so complex. Repetition also uses subjects' time very efficiently, which reduces the already high cash cost of running such experiments. However, repetition also requires Bloomfield and Wilks to balance the orders of the treatments, to ensure that treatment effects are not confounded with order effects.

Tan, Libby, and Hunton (2000) also suggest the use of a combination of between- and within-subjects designs as a method of partitioning the effects of unintentional biases from intentional judgment policies. Following Kahneman and Tversky (1996), they suggest that the between-subjects design provides a clean test of the subject's natural reasoning process, while the within-subjects design draws attention to the independent variable of interest and thus gives the subject a chance to detect and correct errors and inconsistencies in their responses. Comparison of results under the two approaches highlights how subjects address any conflict between what they do and what they know. Evidence of differences using between-subjects treatments, but not using within-subjects treatments, suggests that the between-subjects differences are unintentional. On the other hand, evidence of differences using within-subjects treatments, but not using between-subjects treatments, suggests that subjects are aware of the

implications of the differences in the stimuli, but that, in their natural reasoning process, the stimuli were ignored or subjects' related knowledge was not accessed and used. This method should be useful in other studies that attempt to distinguish between the effects of judgment heuristics versus knowledge.

The choice of between- versus within-subjects designs affects analyses, since within-subjects manipulation (i.e., repeated measures) yields observations that are not independent. For example, Bloomfield and Wilks (2000) observe well over a thousand closing prices in their study. However, since there are only eight distinct groups of subjects, their repeated-measures analyses effectively compute the average treatment effect (a signed difference) for each group, and then perform a t-test on the 8 differences. This design is more powerful than it might seem, because each of the 8 numbers is the average of a large number of observations, and therefore has very little noise.

As discussed in section 4.1.2, most laboratory markets conduct supplementary analyses that break a theory into parts using measured intervening variables. For example, Bloomfield and Wilks (2000) examine how disclosure quality affects market price through its effects on market liquidity, which is measured. It is more difficult to apply pure repeated-measures statistical techniques to such analyses. However, experimenters should be aware that inappropriate statistical methods overstate sample size (and therefore understate p-values), and should be interpreted with caution. More importantly, researchers must make every attempt to use repeated-measures analyses for their main hypothesis tests.

4.2.4. Using Laboratory Financial Markets

When is it necessary to place individuals in laboratory markets? Critics of individual decision-making experiments often suggest that biases and suboptimal behavior would be driven away by market forces. In our view, this criticism alone rarely justifies the cost of a market experiment. As discussed in section 3.3, few experiments have shown that market forces eliminate biases; even when they mitigate a bias, they tend to affect its magnitude, but not its

sign (e.g., market prices are still too high, but not by as much). Because only directional effects are easily generalized from experiments to target settings, using a financial market does not substantially alter an experiment's effectiveness. On the other hand, the market does dramatically increase the cost of the experiment. A group of 50 subjects will yield 50 judgments that are statistically independent of one another. Forming those subjects into ten separate 5-trader markets yields only ten judgments (market prices) that are statistically independent of one another. As a result, the use of a market either reduces power or increases the costs of the study.

Laboratory markets are most appropriate when examining particular forces within the market that might affect bias mitigation (such as the smart-trader hypothesis), or when examining dependent variables that are simply undefined at the individual level (such as trading volume or market liquidity). Even in these cases, however, one can sometimes address experimental goals in individual decision-making tasks. For example, Nelson, Krische, and Bloomfield (2000) use an individual decision-making task to examine how confidence in one's own ability to "pick winners," relative to confidence in large-sample anomalies (such as post-earnings-announcement drift) can affect traders' willingness to rely on a disciplined trading strategy. They do not have traders transact with each other, but rather examine the number of shares that each trader offers to transact. This approach allows researchers to examine the relation between judgment and trading behavior, but does not allow researchers to capture strategic interactions between market participants.

Given that one chooses to conduct a financial market, there are many decisions that can reduce the cost of each observation. One method used almost universally in laboratory financial markets and laboratory games (as in sections 3.3. and 3.4) is to have each group provide many observations (a repeated-measures design). As noted in section 4.2.3, repeated-measures designs offer many advantages, but affect the statistical analyses that must be performed.

5. Conclusions

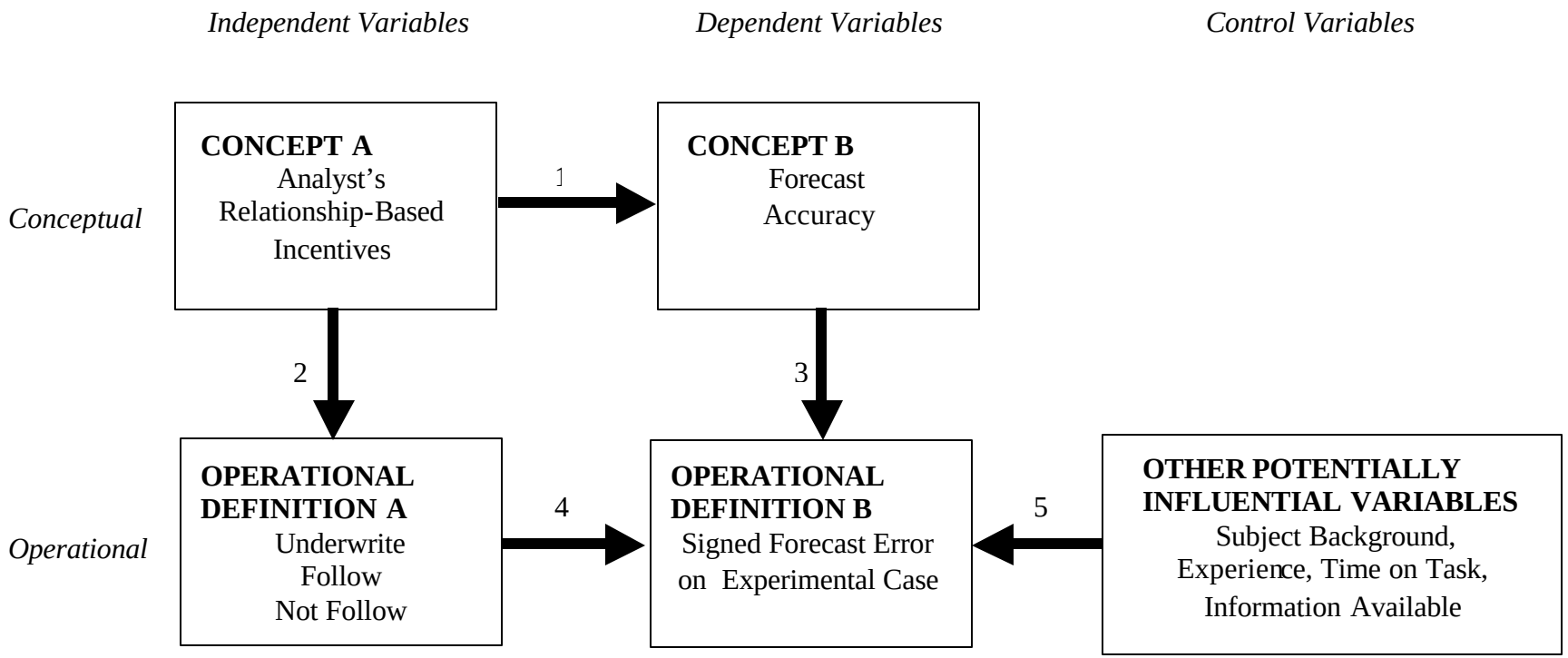
This paper discusses how recent experimental research in financial accounting has responded to past criticisms, discusses how the recent literature has developed and how it can be extended, and provides our perspective on how future experiments can be designed to maximize both effectiveness and efficiency. Our comments are driven by our belief that experiments—whether based on psychological or economic theory—must exploit the primary advantages of the experimental method. Those advantages include the ability to construct an environment in which a causal theory of phenomena can be tested with a maximum of internal validity.

Experimental research is still only a small part of empirical financial accounting research. This raises the question of how financial accounting experiments should relate to the more dominant archival-empirical work. One of the most notable characteristics of the better studies that we have reviewed is their close tie to formal or informal empirical observation. These observations often provide part of the motivation for the experimental studies, and are relied upon to demonstrate the external validity of experimental results.

Future research can relate even more closely to this literature by testing alternative potential explanations for archival findings when there are natural confounds, measurement problems, or where causality is unclear, by explaining contradictory findings, and by examining conditions where large samples are unavailable. Experiments can also point to directions for future archival-empirical studies by specifying either the limits to the generality of existing findings or other findings that should exist in further archival studies.

FIGURE 1

Predictive Validity Framework



BIBLIOGRAPHY

Abarbanell, J.S., and Bernard, V.L. (1992). Tests of Analysts' Overreaction/Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behavior. *Journal of Finance,* 47 (3): 1181-1208.

Abdel-Khalik, A.R., and El-Sheshi, K. (1980) Information Choice and Cue Utilization in an Experiment on Default Prediction. *Journal of Accounting Research,* 18 (2): 325-342.

Ackert, L. F., Church, B. K, and Shehata, M. (1997). An Experimental Examination of the Effects of Forecast Bias on Individuals' Use of Forecasted Information. *Journal of Accounting Research,* 35 (1): 25-42.

Anderson, M.J. (1988). A Comparative Analysis of Information and Evaluation Behavior of Professional and Non-Professional Financial Analysts. *Accounting, Organizations and Society,* 13 (5): 431-446.

Anderson, M. J., and Sunder, S. (1995). Professional Traders as Intuitive Bayesians. *Organizational Behavior and Human Decision Processes,* 64 (2): 185-202.

Andrade, G. (1999). Do Appearances Matter? The Impact of EPS Accretion and Dilution on Stock Prices, Working Paper, Harvard Business School.

Ball, R. (1992). The Earnings-Price Anomaly. *Journal of Accounting and Economics,* 15 (2,3): 319-345.

Ball, R., and Bartov, E. (1996). How Naïve Is the Stock Market's Use of Earnings Information? *Journal of Accounting and Economics,* 21 (3): 319-337.

Bamber, L. (1987). Unexpected Earnings, Firm Size, and Trading Volume Around Quarterly Earnings Announcements. The *Accounting Review,* 62 (3): 510-532.

Bamber, L., Barron, O., and Stober, T. (1997). Trading Volume and Different Aspects of Disagreement Coincident with Earnings Announcements, *The Accounting Review,* 72 (4): 575-597.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A Model of Investor Sentiment. *Journal of Financial Economics,* 49 (3): 307-343.

Bazerman, M. H. (1998). *Judgment in Managerial Decision Making*. New York: Wiley.

Bazerman, M. H., Morgan, K. P., and Loewenstein, G. F. (1997). The Impossibility of Auditor Independence. *Sloan Management Review* (Summer): 89-94.

Beeler, J., and Hunton, J. E. (2001). Contingent Economic Rents: Insidious Threats to Auditor Independence. Working paper, South Florida University.

Beresford, D.R. (1994). A Request for More Research to Support Financial Accounting Standard Setting AAA—Accounting, Behavior and Organization Section. *Behavioral Research in Accounting,* 6 (Supplement): 190-203.

Berg, J., Dickhaut, J., and McCabe, K. (1995). The Individual Versus the Aggregate. In *Judgment and Decision-Making Research in Accounting and Auditing*, Ashton, R. H. and Ashton, A. H. (Eds.) New York: Cambridge.

Bernard, V.L. (1993). Stock Price Reactions to Earnings Announcements: A Summary of Recent Anomalous Evidence and Possible Explanations. In *Advances in Behavioral Finance*, Edited By R. Thaler, 303-340.

Bernard, V.L., and Skinner, D. J. (1996). What Motivates Managers' Choice of Discretionary Accruals? *Journal of Accounting and Economics*, 22 (1-3): 313-325.

Bernard, V.L., and Thomas, J. (1989). Post-Earnings Announcement Drift: Delayed Price Response or Risk Premium? *Journal of Accounting Research*, 27 (1): 1-48.

Bernard, V.L., and Thomas, J. (1990). Evidence That Stock Prices Do Not Fully Reflect the Implications of Current Earnings for Future Earnings. Journal *of Accounting and Economics*, 13 (4): 305-340.

Bernheim, D. (1984). Rationalizable Strategic Behavior. *Econometrica*, 52 (5): 1007-1028.

Bhushan, R. (1994). An Informational Efficiency Perspective on the Post-Earnings-Announcement Drift. *Journal of Accounting and Economics*, 18 (1): 45-65.

Biggs, S.F. (1984). Financial Analysts' Information Search in the Assessment of Corporate Earning Power. *Accounting, Organizations and Society*, 9 (3,4): 313-323.

Biggs, S.F., Bedard, J.C., Gaber, B.G., and Linsmeier, T.J. (1985). The Effects of Task Size and Similarity on the Decision Behavior of Bank Loan Officers. *Management Science*, 31 (8): 970-987.

Bloomfield, R. (1996a). The Interdependence of Reporting Discretion and Informational Efficiency in Laboratory Markets. *The Accounting Review*, 71 (4): 493-511.

Bloomfield, R. (1996b). Quotes, Prices and Estimates of Value in a Laboratory Market. *Journal of Finance*, 51 (5): 1791-1808.

Bloomfield, R. and Hales, J. (2000). Developing Reputations for Reliable Reporting: The Role of Expectations. Working Paper, Cornell University.

Bloomfield, R., and Libby, R. (1996). Market Reactions to Differentially Available Information in the Laboratory. *Journal of Accounting Research*, 34 (2): 183-207.

Bloomfield, R., Libby, R., and Nelson, M. W. (1996). Communication of Confidence as a Determinant of Group Judgment Accuracy. *Organizational Behavior and Human Decision Processes*, 68 (3): 287-300.

Bloomfield, R., Libby, R., and Nelson, M. W. (1999). Confidence and the Welfare of Less-Informed Investors. *Accounting, Organizations and Society*, 24 (8): 623-647.

Bloomfield, R., Libby, R., and Nelson, M. W. (2000a). Over-Reliance on Previous Years' Earnings. Working Paper, Cornell University.

Bloomfield, R., Libby, R., and Nelson, M. W. (2000b). Under-Reactions, Over-Reactions, and Moderated Confidence. *Journal of Financial Markets*, 3: 113-137.

Bloomfield, R. and Wilks, T. J. (2000). Disclosure Effects in the Laboratory: Liquidity, Depth and the Cost of Capital. The *Accounting Review,* 75 (1): 13-42.

Bonner, S. E. (1990) Experience Effects in Auditing: the Role of Task-Specific Knowledge. *The Accounting Review,* 65 (1): 72-92.

Bonner, S. E., and Walker, P. L. (1994). The Effects of Instruction and Experience on the Acquisition of Auditing Knowledge. *The Accounting Review,* 69 (1): 157-178.

Bouwman, M.J. (1984). Expert Vs. Novice Decision Making in Accounting: A Summary. *Accounting, Organizations and Society,* 9 (3,4): 325-327.

Brown, L., and Han, J. (2000). Do Stock Prices Reflect the Implications of Current Earnings for Future Earnings for AR1 Firms? *Journal of Accounting Research,* (Forthcoming).

Calegari, M., and Fargher, N. L. (1997). Evidence That Prices Do Not Fully Reflect the Implications of Current Earnings for Future Earnings: An Experimental Markets Approach. *Contemporary Accounting Research,* 14 (3): 397-433.

Camerer, C. (1987). Do Biases in Probability Judgment Matter in Markets, Experimental Evidence. *American Economic Review,* 77 (5): 981-97.

Camerer, C. (1992). The Rationality of Prices and Volume in Experimental Markets. *Organizational Behavior and Human Decision Processes,* 51 (2): 237-272.

Camerer, C. (1997). Rules for Experimenting in Psychology and Economics, and Why They Differ. In *Understanding Strategic Interaction : Essays in Honor of R Selten*. Van Dam, et al., eds. Berlin ; New York : Springer.

Carroll, J.S., and Johnson, E. (1990). *Decision Research: A Field Guide.* Sage.

Chan, L., Jegadeesh, K. C., and Lakonishok, J. (1996). Momentum Strategies. *Journal of Finance,* 51 (5): 1681-1713.

Clement, M. (1999). Analyst Forecast Accuracy: Do Ability, Resources, and Portfolio Complexity Matter? *Journal of Accounting and Economics,* 27 (3): 285-303.

Cloyd, C.B., Pratt, J., and Stock, T. (1996). The Use of Financial Accounting Choice to Support Aggressive Tax Positions: Public and Private Firms. *Journal of Accounting Research,* 34 (1): 23-43.

Coller, M. (1996). Information, Noise, and Asset Prices: an Experimental Study. *Review of Accounting Studies,* 1: 35-50.

Cuccia, A. D., Hackenbrack, K., and Nelson, M. W. (1995). The Ability of Professional Standards to Mitigate Aggressive Reporting. *The Accounting Review,* 70 (2): 227-248.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor Psychology and Security Market Under- and Overreactions. *Journal of Finance,* 53 (6): 1839-1885.

De Long, J. B., Shleifer, A. Summers, L. H., and Waldmann, R. J. (1991). The Survival of Noise Traders in Financial Markets. *The Journal of Business,* 64 (1): 1-19.

DeBondt, W., and Thaler, R. (1985). Does the Stock Market Overreact. Journal *of Finance,* 40 (3): 793-818.

DeBondt, W., and Thaler, R. (1987). Further Evidence of Investor Overreaction and Stock Market Seasonality. *Journal of Finance*, 42 (3): 557-581.

DeBondt, W., and Thaler, R. (1990). Do Security Analysts Overreact? *American Economic Review,* 80 (2): 52-57.

Dechow, P., and Sloan, R. (1997). Returns to Contrarian Investment Strategies: Tests of Naïve Expectation Hypotheses. *Journal of Financial Economics,* 43 (1): 3-27.

Dechow, P.M., Sloan, R. G., and Sweeney, A. P. (1995). Detecting Earnings Management. *The Accounting Review,* 70 (2): 193-225.

Dietrich, J.R., Kachelmeier, S.J., Kleinmuntz, D. N., and Linsmeier, T. J. (2000). Market Efficiency, Bounded Rationality, and Supplemental Business Reporting Disclosures. *Journal of Accounting Research*, (Forthcoming).

Dopuch, N., and King, R. R. (1996) The Effects of Lowballing on Audit Quality: An Experimental Markets Study. *Journal of Accounting, Auditing and Finance*, 11: 45-69.

Dyckman, T.R. (1964). On the Investment Decision. *The Accounting Review*, 39 (2): 285-295.

Einhorn, H. J. (1980). Learning From Experience and Suboptimal Rules in Decision Making. In T. Wallsten (Ed.) *Cognitive Processes in Choice and Decision Behavior*. Hillsdale, N.J.: Erlbaum.

Einhorn, H. J., and Hogarth, R. M. (1981). Behavioral Decision Theory: Processes of Judgment and Choice. *Annual Review of Psychology,* 32: 53-88.

Einhorn, H. J., and Hogarth, R. M. (1986). Decision Making Under Ambiguity. *Journal of Business,* 59 (4): S225-S250.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance,* 25 (2): 383-417.

Fama, E. F. (1998). Market Efficiency, Long-Term Returns, and Behavioral Finance. *Journal of Financial Economics*, 49 (3): 283-306.

Fischer, P., and Verrecchia, R. (1999). Public Information and Heuristic Trade. *Journal of Accounting and Economics,* 27 (1): 89-124.

Forsythe, R., and Lundholm, R. (1990). Information Aggregation in an Experimental Market. *Econometrica,* 58 (2): 309-48.

Forsythe, R., Lundholm, R., and Reitz, T. (1999). Cheap Talk, Fraud and Adverse Selection in Financial Markets: Some Experimental Evidence. *Review of Financial Studies,* 12: 581-518.

Foster, G., Olsen, C., and Shevlin, T. (1984). Earnings Releases, Anomalies, and the Behavior of Security Returns. *The Accounting Review,* 59 (4): 574-603.

Frankel, R., and Lee, C. (1998). Accounting Valuation, Market Expectation, and Cross-Sectional Stock Returns. *Journal of Accounting and Economics,* 25 (3): 283-319.

Ganguly, A. R., Kagel, J. H., and Moser, D. V. (1994). The Effects of Biases in Probability Judgments on Market Prices. *Accounting, Organizations and Society,* 19 (8): 675-700.

Gervais, S., and Odean, T. (1997). Learning to Be Overconfident, Unpublished Working Paper, University of Pennsylvania.

Ghosh, D., and Whitecotton, S. M. (1997). Some Determinants of Analysts' Forecast Accuracy. *Behavioral Research in Accounting*, 9 (Supplement): 50-68.

Gibbins, M., Salterio, S., and Webb, A. (2000). Evidence About Auditor-Client Management Negotiation Concerning Client's Financial Reporting. *Journal of Accounting Research,* (Forthcoming).

Gibbins, M., and Swieringa, R.J. (1995) Twenty Years of Judgment Research in Accounting and Auditing. In *Judgment and Decision-Making Research in Accounting and Auditing*, Ashton, R. H. and Ashton, A. H. (Eds.) New York: Cambridge.

Gillette, A. B., Stevens, D. E., Watts, S. G., and Williams, A. W. (1999). Price and Volume Reactions to Public Information Releases: an Experimental Approach Incorporating Traders' Subjective Beliefs. *Contemporary Accounting Research*, 16 (3): 437-479.

Gode, D., and Sunder, S. (1993). Allocative Efficiency of Markets With Zero-Intelligence Traders: Market as A Partial Substitute for Individual Rationality. *The Journal of Political Economy.* 101 (1) (February): 119-140.

Gode, D., and Sunder, S. (1997). What Makes Markets Allocationally Efficient? *The Quarterly Journal of Economics*, 112 (2): 603-630.

Gonedes, N., and Dopuch, N. (1974). Capital Market Equilibrium, Information Production, and Selecting Accounting Techniques: Theoretical Framework and Review of Empirical Work. *Journal of Accounting Research,* 12 (Supplement): 48-129.

Griffin, D., and Tversky, A. (1992). The Weighing of Evidence and the Determinants of Confidence. *Cognitive Psychology,* 24 (3): 411-435.

Hackenbrack, K., and Nelson, M. W. (1996). Auditors' Incentives and Their Application of Financial Accounting Standards. *The Accounting Review*, 71 (1): 43-59.

Hand, J. (1990). A Test of the Extended Functional Fixation Hypothesis. *The Accounting Review,* 65 (4): 740-763.

Haynes, C.M., and Kachelmeier, S.J. (1998). The Effects of Accounting Contexts on Accounting Decisions: A Synthesis of Cognitive and Economic Perspectives in Accounting Experimentation. *Journal of Accounting Literature*, 17: 97-136.

Healy, P. M., and Wahlen, J. M. (1999). A Review of the Earnings Management Literature and Its Implications for Standard Setting. *Accounting Horizons,* 13 (4): 365-383.

Herrnstein, R., and Vaughn, W. (1980). Melioration and Behavioral Allocation. In *Limits to Action: the Allocation of Individual Behavior*. Edited By J. Staddon. New York, NY. Academic Press.

Hirst, D. E. (1994). Auditor Sensitivity to Earnings Management. *Contemporary Accounting Research*, 11 (1): 405-422.

Hirst, D. E., and Hopkins, P. E. (1998). Comprehensive Income Reporting and Analysts' Valuation Judgments. *Journal of Accounting Research,* 36 (Supplement): 47-75.

Hirst, D. E., Koonce, L., and Miller, J. (1999). The Joint Effect of Management's Prior Forecast Accuracy and the Form of Its Financial Forecasts on Investor Judgment. *Journal of Accounting Research*, 37 (Supplement): 101-124.

Hirst, D. E., Koonce, L., and Simko, P. J. (1995). Investor Reactions to Financial Analysts' Research Reports. *Journal of Accounting Research*, 33 (2): 335-351.

Hodder, L., L. Koonce and M. L McAnally (2001). SEC Market Risk Disclosures: Implications for Judgment and Decision Making. *Accounting Horizons* (forthcoming).

Hogarth, R. M. (1993). Accounting for Decisions and Decisions for Accounting. *Accounting, Organizations and Society*, 18 (5): 407-424.

Hogarth, R. M. and Einhorn, H.J. (1992). Order Affects in Belief Updating: the Belief-Adjustment Model. *Cognitive Psychology*, 24 (1): 1-55.

Hopkins, P. E. (1996). The Effect of Financial Statement Classification of Hybrid Financial Instruments on Financial Analysts' Stock Price Judgments. *Journal of Accounting Research*, 34 (Supplement): 33-50.

Hopkins, P. E., Houston, R.W., and Peters, M. F. (2000). Purchase, Pooling, and Equity Analysts' Valuation Judgments. The Accounting Review, 75 (3): 257-281.

Hunton, J. E., and McEwen, R. A. (1997). An Assessment of the Relation Between Analysts' Earnings Forecast Accuracy, Motivational Incentives and Cognitive Information Search Strategy. *The Accounting Review*, 72 (4): 497-515.

Jacob, J., Lys, T., and Neale, M. (1999). Expertise in Forecasting Performance of Security Analysts. *Journal of Accounting and Economics*, 28 (1): 51-82.

Jensen, R. (1966). An Experimental Design for Study of Effects of Accounting Variations in Decision Making. *Journal of Accounting Research*, 4 (2): 224-238.

Jung, W., and Kwon, Y. (1988). Disclosure When the Market Is Unsure of Information Endowment of Managers. *Journal of Accounting Research*, 26 (1): 146-153.

Kachelmeier, S. (1996a). Do Cosmetic Reporting Variations Affect Market Behavior? A Laboratory Study of the Accounting Emphasis on Unavoidable Costs. *Review of Accounting Studies*, 1: 115-140.

Kachelmeier, S. (1996b). Discussion of "Tax Advice and Reporting Under Uncertainty: Theory and Experimental Evidence." *Contemporary Accounting Research*, 13: 81-90.

Kachelmeier, S., and Shehata, M. (1992). Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence From the People's Republic of China. *The American Economic Review*, 82: 1120-1141.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: an Analysis of Decision Under Risk. *Econometrica*, 47 (2): 263-291.

Kahneman, D., and Tversky, A. (1996). On the Reality of Cognitive Illusions. *Psychological Review*, 103 (3): 582-588.

Kennedy, J., Kleinmuntz, D.N., and Peecher, M.E. (1997). Determinants of the Justifiability of Performance in Ill-Structured Audit Tasks. *Journal of Accounting Research*, 35 (Supplement): 105-123.

Kennedy, J., Mitchell, T., and Sefcik, S. E. (1998). Disclosure of Contingent Environmental Liabilities: Some Unintended Consequences? *Journal of Accounting Research*, 36 (Autumn): 257-277.

Kim, O., and Verrecchia, R. (1994). Market Liquidity and Volume Around Earnings Announcements. *Journal of Accounting and Economics*, 17 (1,2): 41-67.

King, R. R. (1996). Reputation Formation for Reliable Reporting: an Experimental Investigation. The *Accounting Review*, 71 (3): 375-396.

King, R. R. (2001), An experimental investigation of self-serving biases in an auditing trust game: The effect of group affiliation. Working Paper, Washington University.

King, R. R., and Wallin, D. E. (1991a). Market-Induced Information Disclosures: An Experimental Markets Investigation. *Contemporary Accounting Research*, 8 (1): 170-197.

King, R. R., and Wallin, D. E. (1991b). Voluntary Disclosures When Seller's Level of Information Is Unknown. *Journal of Accounting Research*, 29 (1): 96-108.

King, R. R., and Wallin, D. E. (1995). Experimental Tests of Disclosure With an Opponent. *Journal of Accounting and Economics*, 19 (1): 139-168.

Kinney, W. R. (1986). Empirical Accounting Research Design for Ph.D. Students. The *Accounting Review*, 61 (2): 338-350.

Kinney, W. R., and Martin, R. D. (1994). Does Auditing Reduce Bias in Financial Reporting? A Review of Audit-Related Adjustment Studies. *Auditing: A Journal of Practice and Theory*, 13 (1): 149-156.

Kinney, W. R., and Nelson, M. W. (1996). Outcome Information and the 'Expectations Gap': The Case of Loss Contingencies. *Journal of Accounting Research*, 34 (2): 281-299.

Kothari, S.P. (2000). Capital Markets Research in Accounting. *Journal of Accounting and Economics*, (Forthcoming).

Kunda, Z. 1990. The case for motivated reasoning. *Psychological Bulletin* 108 (3): 480-498.

Kyle, A. S., and Wang, F.A. (1997). Speculation Duopoly With Agreement to Disagree, Can Overconfidence Survive the Market Test? *Journal of Finance*, 52 (5): 2073-2090.

LaPorta, R. (1996). Expectations and the Cross-Section of Stock Returns. *Journal of Finance*, 51 (5): 1715-1742. Lee, C., and Swaminathan, B. (2000). Price Momentum and Trading Volume. *Journal of Finance*, (Forthcoming).

Lee, C., Myers, J., Swaminathan, B. (1999). What Is the Intrinsic Value of the Dow? *Journal of Finance*, 54 (5): 1693-1741.

Lee, C., and Swaminathan, B. (2000). Price Momentum and Trading Volume. *Journal of Finance*, (Forthcoming).

Libby, R. (1981). *Accounting and Human Information Processing: Theory and Applications.* Englewood Cliffs: Prentice-Hall.

Libby, R., and Kinney, W. R. (2000). Earnings Management, Audit Differences, and Analysts' Forecasts. *The Accounting Review*, (Forthcoming).

Libby, R., and Luft, J. (1993). Determinants of Judgment Performance in Accounting Settings: Ability, Knowledge, Motivation, and Environment. *Accounting, Organizations and Society*, 18 (5): 425-450.

Libby, R., and Tan, H-T. (1999). Analysts' Reactions to Warnings of Negative Earnings Surprises. *Journal of Accounting Research*, 37 (2): 415-436.

Lipe, M. G. (1991) Counterfactual Reasoning as a Framework for Attribution Theories. *Psychological Bulletin*, 109 (3): 456-471.

Lipe, M. G. (1998). Individual Investors' Risk Judgments and Investment Decisions: the Impact of Accounting and Market Data. *Accounting, Organizations and Society*, 23 (7): 625-640.

Lundholm, R. J. (1991). What Affects the Efficiency of A Market? Some Answers From the Laboratory. The *Accounting Review*, 66 (3): 486-515.

Maines, L. A. (1994). The Role of Behavioral Accounting Research in Financial Accounting Standard Setting. *Behavioral Research in Accounting*, 6 (Supplement): 204-212.

Maines, L. A. (1995). Judgment and Decision-Making Research in Financial Accounting: A Review and Analysis. In *Judgment and Decision-Making Research in Accounting and Auditing*, Ashton, R. H. and Ashton, A. H. (Eds.) New York: Cambridge.

Maines, L. A., and Hand, J.R.M. (1996). Individuals' Perceptions and Misperceptions of Time Series Properties of Quarterly Earnings. *The Accounting Review*, 71 (3): 317-336.

Maines, L. A., Mautz, R. D., Wright, G. B., Graham, L. E., Rosman, A. J., and Yardley, J. A. (2000). Implications of International Diversity in Joint Venture Financial-Reporting Standards for Financial Analysts' Stock Values. Working Paper, Indiana University.

Maines, L. A., and McDaniel, L. S. (2000). Effects of Comprehensive Income Volatility on Nonprofessional Investors' Judgments: the Role of Presentation Format. *The Accounting Review*, (Forthcoming).

Maines, L. A., McDaniel, L. S., and Harris, M. S. (1997). Implications of Proposed Segment Reporting Standards for Financial Analysts' Investment Decisions. *Journal of Accounting Research*, 35 (Supplement): 1-24.

Mayhew, B. W., Schatzberg, J. W., and Sevcik, G. R. (2000). The Effect of Accounting Uncertainty and Auditor Reputation on Auditor Independence. Working Paper, University of Wisconsin – Madison.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*, Cambridge University Press, Cambridge, UK.

Mear, R., and Firth, M. (1987). Cue Usage and Self-Insight of Financial Analysts. *The Accounting Review*, 62 (1): 176-182.

Mikhail, M., Walther, B. and Willis, R. (1997). Do Security Analysts Improve Their Performance With Experience? *Journal of Accounting Research*, 35 (Supplement): 131-157.

Milgrom, P., and Stokey, N. (1982). Information, Trade and Common Knowledge. *Journal of Economic Theory*, 26 (1): 17-27.

Moser, D. V. (1998). Using an Experimental Economics Approach in Behavioral Accounting Research. *Behavioral Research in Accounting,* 10 (Supplement): 94-110.

Nelson, M. W., Elliott, J. A., and Tarpley, R. L. (2000). Where Do Companies Attempt Earnings Management, and When Do Auditors Prevent It? Working Paper, Cornell University.

Nelson, M. W., and Kinney, W. R. (1997). The Effect of Ambiguity on Auditors' Loss Contingency Reporting Judgments. *The Accounting Review,* 72 (2): 257-274.

Nelson, M. W., Krische, S. D., and Bloomfield, R. (2000). Sticking With the Program: Why Investors Don't Exploit Anomalies Shown in Large-Sample Studies. Working Paper, Cornell University.

O'Brien, J., and Srivastava, S. (1991). Dynamic Stock Markets With Multiple Assets: An Experimental Analysis. *Journal of Finance,* 46 (5): 1811-1838.

Odean, T. (1998). Volume, Volatility, Price, and Profit When All Traders Are Above Average. *Journal of Finance,* 53 (6): 1887-1934.

Ou, J., and Penman, S. (1989). Financial Statement Analysis and the Prediction of Stock Returns. *Journal of Accounting and Economics,* 11 (4): 295-329.

Pankoff, L.D., and Virgil, R.L. (1970). Some Preliminary Findings From A Laboratory Experiment on the Usefulness of Financial Accounting Information to Security Analysts. *Journal of Accounting Research,* 8 (Supplement): 1-48.

Paquette, L., and Kida, T. (1988). The Effect of Decision Strategy and Task Complexity on Decision Performance. *Organizational Behavior and Human Decision Processes,* 41 (1): 128-142.

Payne, J.W., Bettman, J.R., and Johnson, E.J. (1992). Behavioral Decision Research: A Constructive Processing Perspective. *Annual Review of Psychology,* 43: 87-131.

Pearce, D.G. (1984). Rationalizable Strategic Behavior and the Problem of Perfection. *Econometrica,* 52 (5): 1029-1050.

Phillips, F. (1999). Auditor Attention to and Judgments of Aggressive Financial Reporting. *Journal of Accounting Research,* 37 (1): 167-189.

Plott, C., and Sunder, S. (1988). Rational Expectations and the Aggregation of Diverse Information in Laboratory Security Markets. *Econometrica,* 56 (5): 1085-1118.

Runkel, P., and McGrath. J. (1972). *Research on Human Behavior: A Systematic Guide to Method*. New York: Holt, Rinehart and Winston, Inc.

Salterio, S., and Koonce, L. (1997). The Persuasiveness of Audit Evidence: the Case of Accounting Policy Decisions. *Accounting, Organizations and Society,* 22 (6): 573-587.

Simon, H.A. (1957). *Models of Man*. New York: Wiley.

Sloan, R. (1996). Do Stock Prices Fully Reflect Information in Accruals and Cash Flows About Future Earnings. *The Accounting Review,* 71 (3): 289-315.

Slovic, P., Fleissner, D., and Bauman, W. S. (1972). Analyzing the Use of Information in Investment Decision Making: A Methodological Perspective. *Journal of Business,* 45: 283-301.

Slovic, P., and Lichtenstein, S. C. (1968). The Relative Importance of Probabilities and Payoffs in Risk Taking. *Journal of Experimental Psychology Monograph Supplement,* 78.

Slovic, P., and Lichtenstein, S. C. (1971). Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment. *Organizational Behavior and Human Performance,* 6: 649-744.

Smith, E.E., and Medin, D.L. (1981). *Categories and Concepts.* Cambridge: Harvard.

Smith, V. (1976). Experimental Economics: Induced Value Theory. *American Economic Review,* 66: 274-279.

Swaminathan, B., and Lee, C. (2000). Do Stock Prices Overreact to Earnings News? Working Paper, Cornell University.

Tan, H.T., Libby, R., and Hunton, J. (2000). Analysts' Reactions to Earnings Preannouncement Strategies. Working Paper, Cornell University.

Tan, T.C., and Werlang, S. (1988). The Bayesian Foundations of Solution Concepts of Games. *Journal of Economic Theory,* 45 (2): 379-391.

Tetlock, P. 1992. The impact of accountability on judgment and choice: Toward a social contingency model. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology 25: 331-376.

Thaler, R.H. (1999). The End of Behavioral Finance. Financial Analysts' Journal, 55 (Nov./Dec.): 12-17.

Trotman, K. T. (1996). *Research Methods for Judgment and Decision Making Studies in Auditing*. Melbourne, Australia: Coopers and Lybrand.

Tucker, R. R. (1997). The Relationship Between Public and Private Information: An Experimental Markets Study. *Behavioral Research in Accounting,* 9: 219-249.

Tuttle, B., Coller, M., and Burton, F. G. (1997). An Examination of Market Efficiency: Information Order Effects in A Laboratory Market. *Accounting, Organizations and Society,* 22 (1): 89-103.

Tversky, A., and Kahneman, D. (1974) Judgment Under Uncertainty: Heuristics and Biases. *Science,* 185: 1124-1131.

Vincent, L. (1997). Equity Valuation Implications of Purchase Versus Pooling Accounting. *The Journal of Financial Statement Analysis,* 2 (4): 5-19.

Wagenhofer, A. (1990). Voluntary Disclosure With A Strategic Opponent. *Journal of Accounting and Economics,* 12 (4): 341-363.

Watts, R., and Zimmerman, J. (1986). *Positive Accounting Research*. Prentice Hall. Englewood Cliffs, NJ.

Whitecotton, S.M. (1996). The Effects of Experience and Confidence on Decision Aid Reliance: A Causal Model. *Behavioral Research in Accounting,* 8: 194-216.

Wilks, J. (2001). Predecisional Distortion of Evidence as A Consequence of Real-Time Audit Review. Working Paper, Brigham Young University.

Wright, W.F. (1977). Financial Information Processing Models: an Empirical Study. *The Accounting Review*, 52 (3): 676-689.

Yetton, P. W., & Bottger, P. C. (1982). Individual versus group problem solving: An empirical test of a best-member strategy. *Organizational Behavior and Human Decision Processes*, 29: 307-321.