MANAGEMENT SCIENCE
Vol. 52, No. 4, April 2006, pp. 489–500
ISSN 0025-1909 | EISSN 1526-5501 | 06 | 5204 | 0489

DOI 10.1287/mnsc.1050.0485

Does Past Success Lead Analysts to Become Overconfident?

Gilles Hilary
Department of Accounting, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, acgh@ust.hk

Lior Menzly
Vega Asset Management, 375 Park Avenue, Suite 29, New York, New York 10152-0002, lmenzly@vamusa.com

This paper provides evidence that analysts who have predicted earnings more accurately than the median analyst in the previous four quarters tend to be simultaneously less accurate *and* further from the consensus forecast in their subsequent earnings prediction. This phenomenon is economically and statistically meaningful. The results are robust to different estimation techniques and different control variables. Our findings are consistent with an attribution bias that leads analysts who have experienced a short-lived success to become overconfident in their ability to forecast future earnings.

Key words: overconfidence; cognitive biases; analysts; earnings forecasts
History: Accepted by Detlof von Winterfeldt, decision analysis; received July 27, 2004. This paper was with the authors $1\frac{1}{2}$ months for 1 revision.

1. Introduction

This paper studies the short-term dynamics behind analysts' forecasts. We find that analysts who forecast earnings more accurately than the median analyst in the previous four quarters tend to be relatively less accurate and further from the consensus forecast in their subsequent earnings prediction. This result is consistent with analysts becoming overconfident in their ability to forecast future earnings after experiencing a short series of successful predictions.

Analysts are one of the key pillars of financial markets. Further, researchers often use analysts' forecasts as proxies for market expectations and differences in opinion. In addition, analysts' forecasts are one of the rare settings for which researchers have a large natural data set of individuals' actual decisions. Not surprisingly, the activities of analysts have been a fertile ground for behavioral research. In particular, prior literature suggests that analysts (1) make upwardly biased forecasts (e.g., DeBondt and Thaler 1990), (2) overreact to positive information, and (3) underreact to negative information (e.g., Easterwood and Nutt 1999). Yet, these findings have been challenged on several grounds. For example, Gu and Wu (2003) claim that some apparent irrationality in analysts' behavior may be due to skewed data. Whether analysts' forecasts are significantly affected by cognitive biases remains an important but unresolved question.

In this paper, we study from a new perspective the accuracy of individual predictions and its relation to deviations from consensus. Existing empirical literature has studied these two characteristics, but not in a unified framework. In addition, we focus on short-term cyclical variations whereas past research has mainly considered either analysts' fixed characteristics or long-term trends. For example, Sinha et al. (1997) find systematic differences in forecast accuracy across analysts. Yet, they consider analysts' skills as a fixed parameter that varies cross-sectionally but not over time. Other empirical studies have considered the evolution of analysts' characteristics over time. One stream of literature (e.g., Mikhail et al. 1997, Jacob et al. 1999) looks at the effect of experience on forecast accuracy and finds mixed results. Another stream considers the effect of experience on "herding" (i.e., the willingness to deviate from the consensus). Hong et al. (2000) find that analysts deviate more from the consensus as they gain experience. Phillips and Zuckerman (2001) find that analysts are more likely to conform to the norm in the middle of their careers. We depart from these papers by focusing on short-term dynamics rather than long-term trends.[1]

Our results indicate that after making a short series of accurate predictions, analysts are more likely to be inaccurate than their skill and environment would predict. They also take additional risk by deviating from the consensus forecast on their subsequent prediction. This phenomenon is both statistically and economically meaningful. The results are robust to the use of different econometric techniques and to numerous control variables.

[1] One exception, though in a different context, is Richardson et al. (2004). They consider the evolution of accuracy over the 12 months preceding annual earnings announcements.

These joint results are consistent with our research hypothesis that analysts are subject to overconfidence. An individual is said to exhibit overconfidence if he or she overestimates the precision of his or her own information relative to public signals. Prior results indicate that overconfident subjects put too much emphasis on their private information (see Kraemer et al. 2006 for an experimental study and Barber and Odean 2001 for an example of a large-sample test). The combination of the two cognitive principles yields the dynamic notion of overconfidence. In our setting, after analysts have made a series of good predictions, they become overconfident in their ability to predict future earnings. This leads them to put excessive weight on their private information and to discount public signals, such as market reactions and other analysts' forecasts. As a consequence, their next prediction is likely to be more inaccurate and to deviate more from the consensus compared to what it would have been without this cognitive bias. This reduces the likelihood for their next forecast to be superior to the ones from other analysts and, in turn, the prediction may trigger a negative feedback mechanism that reduces overconfidence. Thus, this short-term phenomenon may recurrently appear and disappear. It follows a cyclical pattern in which the intensity varies with the analyst's performance. Random past success creates overconfidence, which, then increases the probability of poor subsequent forecasts and thereby reduces overconfidence. A similar analytical framework has been suggested by Gervais and Odean (2001) but, to our knowledge, no large-sample empirical research has been conducted to investigate this question. Note, however, that analysts acting under this form of overconfidence do not necessarily underperform relative to other analysts but rather they underperform compared to their own expected performance. If the effect of overconfidence is small relative to other characteristics such as skill, it is possible that overconfident analysts consistently outperform analysts who do not suffer from this bias. The analyst would remain "locked" in part of the cycle and permanently exhibit some degree of overconfidence.

This paper contributes to the literature in three ways. First, it investigates whether dynamic overconfidence, a consequence of two major behavioral principles, affects the decision-making process of individuals in an important economic setting. This large sample test complements previous research that was largely based on small sample experimental work. Although the approach focuses on analyst forecasts to take advantage of a particularly rich data set, we expect the results to generalize to other settings as well. Second, the paper also provides new understanding of analyst behavior. Departing from prior analyst literature, we examine the short-term dynamics behind analyst forecasts by considering the influence of past success on current forecasts. Although theoretical papers have considered this issue, to the best of our knowledge, it has not been tested empirically. Previous empirical research, instead, mainly considered either static biases or long-term evolution due, for example, to experience. We contribute to the literature by documenting a phenomenon that is distinct from known biases and robust to extant methodological criticisms. Third, the results suggest a counterintuitive implication for practitioners in the financial markets. If two analysts are believed to possess identical skills and experience but only one of them had a recent series of superior predictions, investors may want to rely more on the subsequent forecast of the less-accurate analyst. In other words, investors may want to downplay the predictions of analysts who have experienced short-term success.

The rest of this paper is organized as follows. In the next section, we discuss the theoretical foundations of our analysis and develop the research hypothesis. In §3, we present the empirical design followed by the estimation in §4. We explore alternative interpretations of the findings in §5. We conclude in §6.

2. Description of the Framework

In this section, we present a theoretical motivation for studying the short-term dynamics of forecasts. We first describe the underlying theoretical foundations using two main principles, self-attribution and overconfidence, before describing their interaction in a unified framework.

2.1. Theoretical Foundations

In a review article, Kunda (1990, p. 480) notes that "there is considerable evidence that people are more likely to arrive at the conclusions that they want to arrive at, but their ability to do so is constrained by their ability to construct seemingly reasonable justifications for these conclusions." Two psychological principles (self-serving attribution and "static" overconfidence) may lead them to do so.

Self-Serving Attribution. Research indicates that individuals employ different causal explanations to account for their successes and failures (e.g., Fitch 1970, Weiner and Kukla 1970, Kukla 1972). According to the self-attribution theory, individuals too strongly attribute events that confirm the validity of their own actions to their ability while attributing events that disconfirm their actions to external noise (Hastorf et al. 1970). For example, Johnson et al. (1964) and Beckman (1970) show that teachers tend to claim responsibility for a student's performance when there is improvement but attribute the blame to various external causes, such as the child's motivation or situational factors, when students perform poorly. Miller (1976) finds that tendencies toward self-attribution are stronger when the task is important ("ego-involving") for the subject.

Overconfidence. A conventional definition for "static" overconfidence implies either extreme beliefs relative to some objective standard (e.g., estimating 90% probability for events that occur less often) or confidence intervals that are too tight (e.g., setting 90% confidence intervals such that "surprises" occur more than 10% of the time). Klayman et al. (1999, p. 216) note that "many studies have reported that the confidence people have in their judgments exceeds their accuracy and that overconfidence increases with the difficulty of the task."[2] The literature offers two main categories of explanations for overconfidence: (a) biases in information processes, and (b) effects of unbiased judgmental errors. The first explanation hypothesizes that individuals search their memories for relevant information and reach a preliminary conclusion. They then proceed to search selectively for more evidence confirming this initial conclusion. The second explanation stresses the possible role of unbiased judgmental error in producing overconfidence. For example, Erev et al. (1994) show that overt responses representing perfectly calibrated true judgments perturbed by random errors can replicate typical patterns observed in empirical studies. Koehler et al. (2002, p. 713), however, remark that "it is not entirely clear what it means to argue that a judge's confidence assessments are not systematically biased but instead merely fail to account for the uncertainty associated with a prediction based on the actual evidence."

A variant of overconfidence applies to the use of private vs. public information. In this setting, overconfident subjects believe that their private information is more accurate than it is and hence "put too much weight on their private information" (Kraemer et al. 2006, p. 424). The existence of this form of overconfidence is consistent with extant experimental literature. Hung and Plott (2001) report that participants in their experimental setting put too much weight on their free private information. In a similar setting, Huck and Oechssler (2000, p. 661) report that the heuristic "follow your (private) signal" explains the observed behavior better than does Bayes' law. In a setting where subjects incurred a cost to acquire private information, Kraemer et al. (2006, p. 424) conclude that "about one-half of the individuals act rationally, whereas the other participants overestimate the private signal value." Bloomfield et al. (2000) report experimental evidence that people are overconfident in their ability to interpret data (relative to the ability of a disciplined trading strategy) and underperform as a result. Larrick and Soll (2004) indicate people tend to combine multiple signals in an inefficient manner by choosing what is perceived as the most informative instead of averaging them. The literature (e.g., Harvey et al. 2000, Yaniv 2004) has also showed that people do not weight advice optimally. For example, Yaniv (2004) reports that people place a higher weight on their own opinion than on an adviser's opinion, even though the use of advice improved accuracy. In addition, more knowledgeable individuals discounted the advice more.

2.2. Framework

The combination of the two cognitive biases yields a dynamic notion of overconfidence and a framework in which an analyst becomes overconfident in her ability to predict future earnings after a series of good predictions. This model is germane to the one developed by Gervais and Odean (2001) or Daniel et al. (1998).[3] The self-attribution principle predicts that analysts who have successfully forecasted earnings in previous periods attribute too much of their success to superior ability and too little of it to chance. The resulting overconfidence in their abilities yields nonoptimal behavior, whereby analysts overweight their private information and rely less on public signals, such as other analysts' forecasts. As a consequence, their next forecast is more likely to be simultaneously further from the consensus forecast and more inaccurate (compared to what it would have been without this cognitive bias). This reduces the likelihood that their next forecast will be superior to the forecasts of other analysts. This, in turn, may trigger a negative feedback mechanism. This inferior performance (relative to the performance expected by the analyst) leads her to revise downward her prior distribution of her skills, which in turn reduces her overconfidence in her skill. Overconfidence, in this case, is not a fixed characteristic but rather a recurring phenomenon whose intensity is dynamic in nature. Under this notion, overconfident analysts do not necessarily underperform other analysts unconditionally but rather they underperform compared to their expected performance if they did not suffer from any cognitive biases. In fact, if the effect of overconfidence is smaller than the effect from other factors such as skill, it is possible that overconfident analysts consistently outperform analysts who do not suffer from this bias and thus remain "locked" in part of the cycle.

[2] Klayman et al. (1999) discuss methodological issues with the initial research but show that, even after controlling for these issues, overconfidence still appears to exist in an experimental setting.

[3] In their model, Daniel et al. (1998) propose a theory of securities-market under- and overreactions based on investors' confidence about the precision of their private information and biased self-attribution causing symmetric shifts in investors' confidence as a function of their investment outcomes.

3. Empirical Design

3.1. Data

The analyst forecast data are retrieved from the Zacks database and cover the period from the last quarter of 1980 to the last quarter of 1997. To increase the consistency of the data, we focus on quarterly predictions made one quarter ahead. We only include the first forecast of the analyst for the company in the quarter. Following Francis and Philbrick (1993, among others), we delete firms that do not have a December year-end report. If an analyst makes more than one forecast in a given quarter, only the first one is considered because the later predictions may be drawn from a different distribution. To obtain a meaningful measure of the consensus, we also require that at least four analysts cover the firm in a particular quarter (e.g., Easterwood and Nutt 1999). To be included in our sample, an analyst needs to have made at least four predictions for a given firm in the preceding quarters. We match the forecast data to the corresponding records of Compustat reported earnings.[4] Abarbanell and Lehavy (2003, p. 105) stress that "a relatively small number of observations can have a disproportionate effect" with noncentral distributions and thus create the appearance of bias. To avoid this issue, we remove outliers by deleting observations in which the difference between the prediction and the realization (scaled by price) is in the top or bottom one percent of the sample. This procedure also mitigates any errors in the data. These data requirements yield a sample of 46,909 quarterly observations.

3.2. Definition of Dependent Variables

Our framework yields two distinct predictions about the effect of short-term past success on both deviation from the consensus and error of the forecast. To test these predictions, we define two dependent variables. *DEV*, the deviation from the consensus, is the absolute value of the difference between the analysts' forecasts and the consensus forecast. We calculate the consensus as the median forecast for a given firm and for the quarter.[5] We interpret *DEV* as a proxy for the likelihood of an analyst to discount public information. *ERROR*, the forecast error, is the absolute value of the difference between actual and forecasted earnings. The realized earnings are computed using either diluted or nondiluted earnings per share (EPS) before extraordinary items (Compustat items 9 or 19), depending on what the analyst targeted in the forecast (as reported in the Zacks database).

3.3. Methodology

In our framework, analysts become more overconfident after a short-term sequence of good predictions. To capture this notion, we define *FREQ* as the number of superior predictions in the preceding four quarters.[6] A superior prediction is a forecast whose error is below the median error in absolute value of all analysts' errors who cover the same firm in that particular quarter. Alternatively, we define *STREAK* as the number of consecutive superior predictions before the current prediction is made.[7] *FREQ* proxies for the frequency of good predictions, whereas *STREAK* proxies for the length of the series of superior predictions. We expect both variables to be significantly and positively associated with both *DEV* and *ERROR*. This specification implicitly assumes that analysts treat each stock independently, assuming this form of mental accounting is not without precedent. For example, Barberis and Huang (2001) propose a model in which investors are loss averse with respect of each stock in their portfolio as opposed to the overall portfolio. This does not mean that there cannot be any feedback between the analysts' performance across different firms. However, we do not have a strong theory suggesting the exact nature of the relation between the different tasks. We therefore prefer to use a simpler model in our main test but we revisit the issue in our robustness tests.

To test our two predictions, we estimate two sets of regressions: one for *DEV* on *FREQ* (alternatively *STREAK*) and the other for *ERROR* on *FREQ* (alternatively *STREAK*) where we control for different potentially confounding factors.

$$\begin{aligned} DEV_t^i = {} & \alpha + \beta FREQ_t^i + \gamma_1 CME_t^i + \gamma_2 CMD_t^i + \gamma_3 T_t^i \\ & + \gamma_4 O_t^i + \gamma_5 \cdot EXPERIENCE_t^i + \gamma_6 LOSS_t^i \\ & + \gamma_7 A_t^i + \gamma_8 MCAP_t^i + \varepsilon_t^i \end{aligned}$$

[4] Both forecasts and realized earnings per share are adjusted on the same basis.

[5] To avoid biases from extreme observations, we use the median instead of the mean. We use all observations available within a quarter to ensure a sufficiently large number of data points to estimate the consensus. The downside of this approach is that it assumes that analysts have complete information of all forecasts that will be made in that quarter. To mitigate the concern, we include control variables that proxy for the amount of information analysts have at the time of the forecast.

[6] We choose four quarters as a compromise between reducing the period too much (which would not let us capture the sequence of the forecasts) and extending the period too much (which would represent the long-run dynamics subsumed by control variables, such as *ACCURACY*). We perform a robustness check on this assumption in §4.5.

[7] Psychologists (e.g., Fiske and Taylor 1991) find that people not only overweight their successes and underweight their failures but also that they overweight successes more than they underweight failures. Therefore, we consider above median performance and not the below median performance. This issue is further studied in §4.4.

$$ERROR_t^i = \alpha + \beta FREQ_t^i + \gamma_1 CME_t^i + \gamma_2 CMD_t^i + \gamma_3 T_t^i + \gamma_4 O_t^i + \gamma_5 \cdot EXPERIENCE_t^i + \gamma_6 LOSS_t^i + \gamma_7 A_t^i + \gamma_8 MCAP_t^i + \varepsilon_t^i,$$

where i is an analyst indicator.

We provide controls for earnings surprises and the level of uncertainty, which affect all analysts covering a given company. Incorporating these controls allows us to focus on the performance of the analyst relative to her peers even though our dependent variables, *ERROR* and *DEV*, are measured in absolute terms. Therefore, the marginal effect of *STREAK* and *FREQ* captures the deviation relative to other analysts. *Current Median Error* (*CME*) is the median error of all analysts who follow a given company for the current quarter. This variable captures earnings surprises and other firm-level shocks. Similarly, *Current Median Deviation* (*CMD*) is the median *DEV* of the analysts who follow a given company for the current quarter. This variable measures the heterogeneity in analyst beliefs about earnings outcomes. We expect *CME* to be positive in the *ERROR* regressions and *CMD* to be positive in the *DEV* regressions. In addition, we employ two variables to proxy for the size of the information set available at the time the prediction is made. *T* (*TIME*) is the number of days between the date of the forecast and the end of the quarter. *O* (*RANK ORDER*) is the number of forecasts published before the analyst issues the prediction divided by the total number of forecasts published over the quarter. We expect variables to be negative in both the *DEV* and *ERROR* regressions. Mikhail et al. (1997) find that experience may lead to improvement in analysts' forecasts. To control for this possible effect, we calculate *EXPERIENCE* as the number of quarters when analysts have made predictions before the current forecast.[8] Based on Mikhail et al. (1997), we expect *EXPERIENCE* to be negative in the *ERROR* regressions. In addition, previous literature (e.g., Abarbanell and Lehavy 2003) finds that the consensus forecast is more upwardly biased and more inaccurate for firms with negative earnings and that losses may explain a considerable portion of previously documented anomalies. To control for this potential effect, we introduce a dummy variable for a loss in the current quarter (*LOSS*) in the regression. We expect *LOSS* to be positive in the *ERROR* regressions. Finally, past literature has shown that firm size is correlated with numerous factors. To proxy for possible correlated omitted variables, we introduce *MCAP* as the natural logarithm of the market value of the firm's equity. For example, *MCAP* could proxy for the amount of information available to the market for a particular firm. We expect *MCAP* to be negative in both the *DEV* and *ERROR* regressions.[9]

We use a panel (fixed-effect) technique to estimate these equations. To do so, we subtract the mean (calculated for each combination of analyst and firm) for each variable before running the regressions. The standard errors in the pooled and fixed-effect regressions are groupwise heteroskedasticity-consistent (i.e., adjusted for clustering by quarter and by firm). This procedure specifies that the observations are independent across groups ("clusters") but not necessarily independent within groups. To mitigate a potential heteroskedasticity problem, we deflate all variables involving *ERROR* and *DEV* by the price at the beginning of the quarter. Hence, all variables except *FREQ*, *STREAK*, *TIME*, and *RANK ORDER* are scaled.[10]

The fixed-effect regression is particularly suitable to test our hypotheses, which focus on the short-term dynamics in analysts' forecasts. It eliminates the cross-sectional variations in the means yet it leaves the time-series dynamics of the analyst forecasts intact. The use of analyst/firm fixed effects provides a natural control for omitted variables. For example, constant differences in analysts' skill levels, prediction biases, or willingness to deviate from the consensus are all controlled for. However, the fixed-effect regressions require estimating numerous parameters because our procedure is equivalent to including dummies for each analyst and forecasted firm combination in the regressions.

To alleviate this burden and to show that our results are robust to different estimation techniques, we also run pooled regressions. Since the skill of the analyst (Jacob et al. 1999) and the analyst's willingness to deviate from the consensus are two potentially important characteristics, we include two additional variables in our cross-sectional regressions. First, we calculate the difference between the analyst error

[8] Results are very similar when we calculate *EXPERIENCE* as the number of quarters when analysts have made predictions for a given firm. Since the database does not extend prior to 1980, there is a mechanical tendency for analysts to be more experienced towards the end of our sample period. To control for this effect, we add either a time trend or a time trend and its interaction with *EXPERIENCE*. The results (not tabulated) are qualitatively similar to the ones reported in the table.

[9] Results are unaffected when controlling for the number of analysts following a firm, the number of firms and sectors followed by an analyst, the number of analysts working for the analyst's employer, or the number of sectors covered by the employer. In addition, analysts making forecasts at the turn of year could be affected differently than those early on during the year, owing to compensation-related issues and because of news about annual reports confirming or disconfirming performance. To control for this possible effect, we add a dummy variable if the quarter ends in December. Our results are not affected.

[10] Similar results are obtained without deflating the variables.

in absolute value and the median error of all analysts for a given quarter for a given company. We define *ACCURACY* (*A*) as the median difference over the entire period of coverage for a given analyst. Since *ACCURACY* (*A*) proxies for the skill of the analyst, we flip the sign of the variable (i.e., multiply the median by minus one) so that the variable is positively correlated with accuracy. We expect *ACCURACY* to be negative in the *ERROR* regressions. Second, we also considered a measure of past deviation from the consensus (the average *DEV* over the last four predictions) in the *DEV* regressions.[11] However, we do not tabulate the results including this second additional control because the variable is consistently nonsignificant and our results are otherwise unchanged. In addition, to ensure that our results are not driven by the inclusion of analyst-specific variables such as *ACCURACY*, we reestimate our cross-sectional regressions controlling only for the variables related to the firms being followed (*MCAP* and *LOSS*). Untabulated results are qualitatively similar to the ones reported. Finally, to provide an alternative adjustment for cross-correlation, we also run Fama-MacBeth (1973) regressions (FM, hereafter). However, since the results for the pooled and FM regressions are qualitatively similar, we do not tabulate the FM ones here.[12]

4. Empirical Results

4.1. Descriptive Statistics

We present general descriptive statistics in Table 1 (Panel A). The mean and median forecasted earnings are greater than the realized ones. This is consistent with previous studies and indicates that forecasts are lower than realized earnings on average. A correlation matrix (Panel B) suggests a positive relation between deviation from consensus (*DEV*) and *ERROR*. The correlation between *FREQ* or *STREAK* and *DEV* or *ERROR* is also positive.[13] In addition, the correlation between the different control variables is reasonably low suggesting that multicollinearity is not an issue. Table 2 stratifies the sample based on the possible values of *FREQ* (Panel A) and *STREAK* (Panel B), and reports key statistics for the dependent variables. As the value of *FREQ* or *STREAK* increases, the mean of *DEV* increases monotonically. A similar pattern exists even after deflating the value by the price at the beginning of the quarter. In addition, the value of *ERROR* also increases with *FREQ* (especially after deflating *ERROR* by price). For example, the mean of *DEV* conditional on the value of *FREQ* increases from 0.032 to 0.087. Results for *STREAK* are similar. A Sheffé test and a Kruskal-Wallis test reject the hypothesis of joint equality of means and median respectively (p-values $= 0.000$). These results are consistent with the proposed dynamics of overconfidence.

4.2. Overconfidence and Deviation from the Consensus

We report the results of the *DEV* regressions using the fixed effects and the pooled specifications in Tables 3 and 4, respectively. In each table, the first column reports the results when *FREQ* is the treatment variable and the second column when *STREAK* is used. The results of the three specifications (fixed effect, pooled, FM) support the overconfidence hypothesis. The coefficients of *FREQ* and *STREAK* are positive and statistically significant. The t-statistics are respectively 3.03 and 2.19 in the pooled regressions. In the fixed-effect regressions, the coefficient on *FREQ* and *STREAK* are significantly positive at the 10% level. This suggests that analysts increase their deviation from the consensus after a series of successes. The coefficients are also economically significant. Analysts whom the present account predicts will be most overconfident (*FREQ* $= 4$) have an expected *DEV* 13% higher than the average deflated *DEV*.[14]

The control variables have the expected signs. The coefficients on *CMD* and *CME* are positive, while *LOSS*, *ACCURACY*, *RANK ORDER* (*O*), *TIME* (*T*), and *MCAP* are negative. *EXPERIENCE* is not significant in the pooled regressions but is significantly negative in the fixed-effect regressions.[15] The R^2 is around 36% in the fixed-effect regressions and it is about 39% in the pooled ones.[16]

[11] We assume zero deviation when lagged values of *DEV* are not defined.

[12] As a robustness check, we run Seemingly Unrelated Regressions (SUR). The statistical significance of our main results is materially improved. Wilks' tests indicate joint significance of *FREQ* or *STREAK* at the 0.01% level. We have also performed a path analysis to investigate whether *ERROR* mediates *DEV* or vice versa. Results indicate that *FREQ* and *STREAK* remain significant after controlling for the possible mediations. We believe this further supports our understanding that overconfidence will lead analysts to jointly make a bolder and more inaccurate prediction.

[13] This is true with a Pearson, Spearman, and Kendall specification. The Spearman and Kendall correlation are based on ranks and do not impose the assumption of linearity.

[14] We multiply the value of the coefficient (0.074 from Table 4) by $(4 - 1.235)$, that is, the maximum value of *FREQ* minus its expected value, and divide the product by the average deflated deviation from consensus (0.0016 from Table 1). The result for *STREAK* is around 19%.

[15] Results using FM regressions are very similar to the pooled results. However, the coefficient on *LOSS* becomes only marginally significant, and the coefficients on *RANK ORDER* (*O*) and *TIME* (*T*) become nonsignificant in the FM regressions.

[16] We also adjust the t-statistics in the FM regressions following the procedure outlined by Fama and French (2000). Results still hold.

Table 1 Descriptive Statistics

Panel A: Descriptive statistics

Variable	Mean	Median	Std. dev.	Q3	Q1
DEV	0.0433	0.0150	0.1584	0.0450	0.0000
DIFLDEV	0.0016	0.0006	0.0057	0.0017	0.0000
ERROR	0.1633	0.0600	0.3340	0.170	0.020
DIFLERR	0.0059	0.0024	0.0103	0.0067	0.0008
FORECAST	0.5263	0.4400	0.5850	0.700	0.250
EARN	0.4906	0.4300	0.6633	0.710	0.220
FREQ	1.2353	1.0000	1.0080	2.000	0.000
STREAK	0.4825	0.0000	0.8632	1.000	0.000

Panel B: Correlation table

	DIFLDEV	*DIFLERR*	*FREQ*	*STREAK*	*CME*	*CMD*	*TIME*	*RANK ORDER*	*EXPERIENCE*	*LOSS*	*MCAP*
DIFLERR	0.19	1.00									
FREQ	0.05	0.04	1.00								
STREAK	0.04	0.03	0.65	1.00							
CME	0.49	0.87	0.05	0.04	1.00						
CMD	0.52	0.23	0.08	0.05	0.37	1.00					
TIME	−0.03	0.02	−0.04	−0.04	−0.01	−0.03	1.00				
RANK ORDER	0.02	−0.02	0.02	0.02	0.01	0.01	−0.76	1.00			
EXPERIENCE	−0.01	0.00	0.02	0.01	0.01	0.00	−0.00	0.01	1.00		
LOSS	0.16	0.55	0.04	0.03	0.53	0.21	−0.01	0.00	0.03	1.00	
MCAP	−0.10	−0.13	0.03	0.02	−0.13	−0.13	0.03	−0.03	0.08	−0.17	1.00
ACCURACY	−0.04	−0.08	0.22	0.16	0.00	−0.01	−0.01	0.03	0.02	−0.00	0.00

Notes. Panel A. *DEV* is the absolute value of the difference between the analyst forecast and the consensus (defined as the median of other analysts' predictions). *ERROR* is the absolute value of the difference between forecasted and realized earnings. *DIFLDEV* and *DIFLERR* are *DEV* and *ERROR* deflated by the price of the stock at the beginning of the quarter. *FORECAST* is the forecast made by the analyst, and *EARN* is the realized earnings per share (Compustat items 19 and 9). *FREQ* is the number of above-median predictions in the last four predictions. *STREAK* is the number of above-median predictions in a row before the current prediction. The mean, standard deviation, median, and all other statistics are computed for the entire sample that begins in the last quarter of 1980 and finishes in the last quarter of 1997. This yields 46,909 observations.

Panel B. *DIFLDEV* and *DIFLERR* are *DEV* and *ERROR* deflated by the price of the stock at the beginning of the quarter. *FREQ* is the number of above-average predictions in the last four predictions. *STREAK* is the number of above-average predictions in a row before the current prediction. *ACCURACY* (*A*) is minus the median (over the entire period of coverage) of the difference between the analyst error and the median error of all analysts for a given quarter and a given company. The correlation estimates are computed for the entire sample that begins in the last quarter of 1980 and finishes in the last quarter of 1997. All correlation estimates are significant with p-values lower than 0.05 (except the correlation of *EXPERIENCE* with either *ERR*, *CME*, *CMD*, *TIME*, *RANK ORDER*, between *ACCURACY* and *MCAP*, between *LOSS* and either *TIME* or *RANK ORDER*, between *ACCURACY* with either *MCAP* or *LOSS*).

4.3. Overconfidence and *ERROR*

Tables 5 and 6 present the results from the fixed-effect and pooled specifications, respectively. Results from all procedures indicate that analysts become less accurate after a series of successful predictions. The coefficients on *FREQ* and *STREAK* are both significantly positive at the 1% level in all specifications. The economic magnitude is such that analysts subject to full overconfidence ($FREQ = 4$) are expected to have 9% higher *ERROR* than the average deflated *ERROR*.[17]

The control variables have the expected signs. The coefficients on *CME* and *LOSS* are positive. The coefficient on *CMD* is negative. The coefficient on *ACCURACY* (*A*) is significantly negative in the pooled regression. The coefficients on *RANK ORDER* (*O*) indicate that earlier predictions are subject to more errors. *TIME* is positive in the pooled regressions. The coefficient on *EXPERIENCE* is negative, but barely significant. This result is generally consistent with Jacob et al. (1999) who find no effect of learning-by-doing.[18, 19] The R^2 reaches 79%, which suggests that the regression succeeds in capturing the expected error in prediction. To control for the magnitude as opposed to the frequency of past success, we add *Past Error* (*PE*, the mean past error over the last four predictions) as an additional variable in untabulated tests. The sign and significance of *FREQ* and

[17] The calculation is similar to the one described in footnote 12 for *DEV*. Analysts subject to full overconfidence are expected to have an *ERROR* higher by 7% in the case of *STREAK*.

[18] Results are qualitatively similar when we use a FM regression. However, the coefficients on *CMD* and *TIME* become nonsignificant.

[19] These results suggest that analysts do not improve over time, but that the inferior ones are "weeded-out" as time passes. Further analysis would be required, however, to verify this point.

Table 2 Statistics of *DEV* and *ERROR* Conditional on the Value of *FREQ*

Panel A: Statistics of *DEV* and *ERROR* conditional on the value of *FREQ*

	FREQ (no. obs)	*DEV*	*ERROR*	*DIFLDEV*	*DIFLERR*
Mean	0	0.032	0.134	0.0013	0.0054
Median	(12,561)	0.030	0.130	0.0013	0.0058
Std		0.095	0.295	0.0039	0.0101
Mean	1	0.039	0.161	0.0015	0.0058
Median	(16,987)	0.040	0.170	0.0016	0.0067
Std		0.126	0.343	0.0036	0.0102
Mean	2	0.053	0.181	0.0019	0.0062
Median	(11,919)	0.050	0.200	0.0020	0.0071
Std		0.214	0.349	0.0076	0.0104
Mean	3	0.058	0.196	0.0020	0.0066
Median	(4,636)	0.060	0.220	0.0022	0.0077
Std		0.133	0.342	0.0058	0.0106
Mean	4	0.087	0.216	0.0030	0.0069
Median	(803)	0.080	0.250	0.0026	0.0083
Std		0.444	0.399	0.0192	0.0105

Panel B: Statistics of *DEV* and *ERROR* conditional on the value of *STREAK*

	STREAK (no. obs)	*DEV*	*ERROR*	*DIFLDEV*	*DIFLERR*
Mean	0	0.040	0.155	0.0015	0.0057
Median	(32,294)	0.040	0.160	0.0016	0.0065
Std		0.147	0.321	0.0046	0.0102
Mean	1	0.047	0.171	0.0017	0.0060
Median	(9,387)	0.050	0.180	0.0018	0.0069
Std		0.163	0.356	0.0070	0.0100
Mean	2	0.054	0.195	0.0019	0.0067
Median	(3,232)	0.050	0.210	0.0021	0.0076
Std		0.126	0.362	0.0041	0.0115
Mean	3	0.060	0.199	0.0021	0.0067
Median	(1,190)	0.065	0.220	0.0023	0.0075
Std		0.129	0.372	0.0054	0.0114
Mean	4	0.087	0.216	0.0030	0.0069
Median	(803)	0.080	0.250	0.0026	0.0083
Std		0.444	0.399	0.0192	0.0105

Notes. The mean, standard deviation, and median are computed for each group for the entire sample that begins in the last quarter of 1980 and finishes in the last quarter of 1997. The descriptive statistics are calculated for groups based on *FREQ* (Panel A) and *STREAK* (Panel B). *FREQ* is the number of above-median predictions in the last four predictions. *STREAK* is the number of above-median predictions in a row before the current prediction. *DEV* is the absolute value of the difference between the analyst forecast and the consensus (defined as the median of other analysts' predictions). *ERROR* is the absolute value of the difference between forecasted and realized earnings. *DIFLDEV* and *DIFLERR* are deviation and error deflated by the price of the stock at the beginning of the quarter.

STREAK are not affected. *PE* is not significant in most regressions, except in the *ERROR* fixed-effect ones, where it is negative as expected.

Although we cannot directly observe the presumed cognitive biases (i.e., self-attribution or overconfidence), the combination of the *DEV* and *ERROR* regressions supports the hypothesis that analysts become overconfident after a series of successful predictions.

Table 3 Fixed-Effect Regressions of *DEV*: Standard Errors Are Groupwise Heteroskedasticity-Consistent

The estimated model:

$$DEV_t^i = \alpha^i + \beta FREQ_t^i + \gamma_1 CME_t^i + \gamma_2 CMD_t^i + \nu_1 T_t^i + \nu_2 O_t^i + \phi_1 \cdot EXPERIENCE_t^i + \psi_1 LOSS_t^i + \psi_2 MCAP_t^i + \varepsilon_t^i$$

Variable	Predicted sign	Dependent variable: *DEV*	Dependent variable: *DEV*
INTERCEPT	(0)	0.000	0.000
		(0.00)	(0.00)
FREQ	(+)	0.065	
		(1.72)	
STREAK	(+)		0.066
			(1.73)
CME	(?)	225.299	225.223
		(2.39)	(2.39)
CMD	(+)	763.429	763.653
		(7.96)	(7.97)
TIME (*T*)	(−)	−0.010	−0.010
		(−2.88)	(−2.88)
RANK ORDER (*O*)	(−)	−0.460	−0.461
		(−2.93)	(−2.93)
EXPERIENCE	(?)	0.013	0.013
		(2.25)	(2.26)
LOSS	(?)	−3.819	−3.818
		(−2.05)	(−2.05)
Log Market Equity (*MCAP*)	(−)	−0.160	−0.159
		(−2.30)	(−2.29)
Number of observations		46,909	46,909
R-square (%)		35.96	35.97

Notes. The dependent variable *DEV* is the absolute value of the difference between the analyst forecast and the consensus (defined as the median of other analysts' predictions). Variables are defined in Table 3. For readability, all coefficients are multiplied by 1,000 in the table. The sample covers 1980 to 1997. To be included, the observation has to be made within the quarter and has to be the first forecast of the analyst for the company in the quarter. For a given company, quarters when fewer than four analysts issue a forecast are excluded. An observation is included if the analyst has made at least four previous predictions on a company. The t-statistics, reported in parentheses, are calculated using (firm and quarter) groupwise heteroskedasticity-consistent standard errors.

4.4. Overconfidence vs. Underconfidence

Psychologists have found that individuals both overweight their successes and underweight their failures. Further, the phenomenon is asymmetric: individuals tend to overweight successes more than they underweight failures. For example, Fiske and Taylor (1991) note that self-enhancing attributions for success are more common than self-protective attributions for failures. Thus, underconfidence should be less significant than overconfidence. To test this, we recompute *STREAK* as the number of quarters that the analyst had an error above the median. The coefficient is negative in both the pooled and in the fixed-effect regressions for both *DEV* and *ERROR*. However, compared with the regressions in which *STREAK* measures superior performances, the coefficient has a

Table 4 Pooled Regressions of *DEV*: Standard Errors Are Groupwise Heteroskedasticity-Consistent

The estimated model:

$$DEV_t^i = \alpha + \beta FREQ_t^i + \gamma_1 CME_t^i + \gamma_2 CMD_t^i + \nu_1 T_t^i + \nu_2 O_t^i + \phi_1 \cdot EXPERIENCE_t^i + \psi_1 LOSS_t^i + \phi_2 \cdot A_t^i + \psi_2 MCAP_t^i + \varepsilon_t^i$$

Variable	Predicted sign	Dependent variable DEV	Dependent variable DEV
INTERCEPT	(?)	0.681 (2.93)	0.724 (3.11)
FREQ	(+)	0.074 (3.03)	
STREAK	(+)		0.085 (2.19)
CME	(?)	216.720 (2.49)	216.717 (2.49)
CMD	(+)	775.433 (7.80)	776.280 (7.82)
TIME (*T*)	(−)	−0.007 (−3.21)	−0.007 (−3.23)
RANK ORDER (*O*)	(−)	−0.224 (−2.09)	−0.227 (−2.12)
EXPERIENCE	(?)	−0.004 (−1.27)	−0.004 (−1.18)
LOSS	(?)	−3.089 (−2.02)	−3.087 (−2.02)
Log Market Equity (*MCAP*)	(−)	−0.076 (−3.71)	−0.075 (−3.68)
ACCURACY (*A*)	(?)	−289.813 (−4.03)	−283.963 (−3.99)
Number of observations		46,909	46,909
R-square (%)		38.83	38.84

Notes. The dependent variable *DEV* is the absolute value of the difference between the analyst forecast and the consensus (defined as the median of other analysts' predictions). The first column includes *FREQ* as the explanatory variable, whereas the second includes *STREAK*. *FREQ* is the number of above-median predictions in the last four predictions. *STREAK* is the number of above-median predictions in a row before the current prediction. *ERROR* is defined as the absolute error of the forecast at time *t* for a given analyst and company. *Current Median ERROR* (*CME*) is the median error of the analysts (excluding the analyst making the forecast) who follow a given company for the current quarter. Similarly, *Current Median DEV* (*CMD*) is the median *DEV* of the analysts (excluding the analyst making the forecast) who follow a given company for the current quarter. *TIME* (*T*) is the number of days between the date of the forecast and the end of the quarter. *RANK ORDER* (*O*) is the ratio of the number of forecasts published before the analyst issues her prediction to the total number of forecasts published over the quarter. *EXPERIENCE* is the number of quarters when analysts have made predictions before they make the current forecast. *LOSS* is an indicator of whether the company reports negative earnings. *ACCURACY* (*A*) is minus one multiplied the median (over the entire period of coverage) of the difference between the analyst error and the median error of all analysts for a given quarter and a given company. Log of Market Equity (*MCAP*) is the natural logarithm of the market value of firm equity. All variables (except *FREQ*, *STREAK*, *TIME*, and *RANK ORDER*) are scaled by the price at the beginning of the quarter. For readability, all coefficients are multiplied by 1,000 in the table. The sample covers 1980 to 1997. To be included, the observation has to be made within the quarter and has to be the first forecast of the analyst for the company in the quarter. For a given company, quarters when fewer than four analysts issue a forecast are excluded. An observation is included if the analyst has made at least four previous predictions on a company. The *t*-statistics, reported in parentheses, are calculated using (firm and quarter) groupwise heteroskedasticity-consistent standard errors.

Table 5 Fixed-Effect Regressions of *ERROR*: Standard Errors Are Groupwise Heteroskedasticity-Consistent

The estimated model:

$$ERROR_t^i = \alpha^i + \beta FREQ_t^i + \gamma_1 CME_t^i + \gamma_2 CMD_t^i + \nu_1 T_t^i + \nu_2 O_t^i + \phi_1 \cdot EXPERIENCE_t^i + \psi_1 LOSS_t^i + \psi_2 MCAP_t^i + \varepsilon_t^i$$

Variable	Predicted sign	Dependent variable ERROR	Dependent variable ERROR
INTERCEPT	(0)	0.000 (0.00)	0.000 (0.00)
FREQ	(+)	0.119 (3.16)	
STREAK	(+)		0.097 (2.58)
CME	(+)	755.261 (8.37)	755.127 (8.38)
CMD	(?)	−521.319 (−3.44)	−520.960 (−3.44)
TIME (*T*)	(−)	0.012 (3.59)	0.012 (3.59)
RANK ORDER (*O*)	(−)	−0.233 (−1.43)	−0.235 (−1.43)
EXPERIENCE	(−)	−0.010 (−1.45)	−0.009 (−1.39)
LOSS	(+)	5.058 (2.83)	5.060 (2.83)
Log Market Equity (*MCAP*)	(−)	−0.665 (−4.72)	−0.664 (−4.71)
Number of observations		46,909	46,909
R-square (%)		76.49	76.49

Notes. The dependent variable *ERROR* is defined as the absolute error of the forecast at time *t* for a given analyst and company. The first column includes *FREQ* as the explanatory variable, whereas the second includes *STREAK*. Variables are defined in Table 3. For readability, all coefficients are multiplied by 1,000 in the table. The sample covers 1980 to 1997. To be included, the observation has to be made within the quarter and has to be the first forecast of the analyst for the company in the quarter. For a given company, quarters when fewer than four analysts issue a forecast are excluded. An observation is included if the analyst has made at least four previous predictions on a company. The *t*-statistics, reported in parentheses, are calculated using (firm and quarter) groupwise heteroskedasticity-consistent standard errors.

smaller magnitude (about two to three times smaller) and a lower significance level (the coefficient ceases to be significant in the fixed-effect regressions) in the *DEV* regressions but is qualitatively similar in the *ERROR* regressions.

4.5. Patterns of Overconfidence Dynamics

We posit that the overconfidence should be a short-term phenomenon. However, its exact length is an empirical question. *FREQ* and *STREAK* are formed based on four lagged predictions. The choice of the length of the period represents a trade-off between obtaining more variations in the *FREQ* and *STREAK* variables at the expense of diluting the effect. Nevertheless, we recompute *FREQ* and *STREAK* using

Table 6 Pooled Regressions of *ERROR*: Standard Errors Are Groupwise Heteroskedasticity-Consistent

The estimated model:

$$ERROR_t^i = \alpha + \beta FREQ_t^i + \gamma_1 CME_t^i + \gamma_2 CMD_t^i + \nu_1 T_t^i + \nu_2 O_t^i + \phi_1 \cdot EXPERIENCE_t^i + \psi_1 LOSS_t^i + \phi_2 \cdot A_t^i + \psi_2 MCAP_t^i + \varepsilon_t^i$$

Variable	Predicted sign	Dependent variable *ERROR*	Dependent variable *ERROR*
INTERCEPT	(?)	2.579	2.726
		(5.02)	(5.05)
FREQ	(+)	0.187	
		(4.20)	
STREAK	(+)		0.123
			(2.64)
CME	(+)	769.593	769.745
		(9.11)	(9.11)
CMD	(?)	−409.192	−405.982
		(−2.17)	(−2.14)
TIME (*T*)	(−)	0.008	0.008
		(3.40)	(3.30)
RANK ORDER (*O*)	(−)	−0.402	−0.415
		(−3.27)	(−3.38)
EXPERIENCE	(−)	−0.006	−0.005
		(−1.82)	(−1.55)
LOSS	(+)	4.805	4.814
		(3.19)	(3.19)
Log Market Equity (*MCAP*)	(−)	−0.172	−0.168
		(−3.80)	(−3.74)
ACCURACY (*A*)	(−)	−1,026.153	−994.266
		(−9.96)	(−9.90)
Number of observations		46,909	46,909
R-square (%)		78.48	78.46

Notes. The dependent variable *ERROR* is defined as the absolute error of the forecast at time t for a given analyst and company. The first column includes *FREQ* as the 00000.explanatory variable whereas the second includes *STREAK*. Variables are defined in Table 3. For readability, all coefficients are multiplied by 1,000 in the table. The sample covers the period 1980 to 1997. An observation is included if the analyst has made at least four previous predictions on a company. The t-statistics, reported in parentheses, are calculated using (firm and quarter) groupwise heteroskedasticity-consistent standard errors.

different numbers of lagged periods (from 2 to 8 quarters). All coefficients for *FREQ* or *STREAK* are positive. In the *ERROR* regressions, *FREQ* is significant in all regressions; *STREAK* is significant up to four quarters. Results in the *DEV* regressions are more sensitive to the length of the period and are typically not significant for shorter horizons.

We also consider the growth and decay of overconfidence. *STREAK* implicitly assumes that overconfidence decay is immediate after one bad performance. *FREQ*, on the other hand, assumes a symmetric pattern in which having one prediction above the median performance has the same effect as having one below the median (with opposite signs). Although the results of the two specifications are quite similar, the magnitude and the significance of the coefficients are slightly larger for *FREQ* than for *STREAK*. This would suggest some persistence in overconfidence.

4.6. Overconfidence and Multitasking

As mentioned in §3.3, we implicitly assume firm-dependent degrees of overconfidence in analysts. To examine whether any carryover effect exists between the predictions made by the analysts across firms, we use the average *FREQ* and *STREAK* per analyst in a given quarter in our regressions. Most results hold. Both means are significant at less than 5% level in the *ERROR* regressions (both cross-sectional and fixed-effect). The mean of *STREAK* is also significant in the *DEV* regression at the 10% level in both the cross-sectional and the fixed-effect regressions. However, the mean of *FREQ* is not significant in the *DEV* regressions. While we are intrigued by the carryover of overconfidence in a multitasking environment, we leave a more rigorous analysis of the interaction between tasks and its effect on overconfidence for future work.

5. Alternative Interpretations of the Results

Our research question is motivated by the importance of psychological biases in shaping human behavior. However, we cannot completely reject rationality since its existence is always conditional on the objective function of the analysts. Our results can therefore be explained by making ad hoc assumptions about the objective function of the analysts. Yet, our tests limit the functions that are consistent with rationality. If it is increasing in accuracy (i.e., analysts are rewarded for being accurate) and in conditional deviation from consensus (i.e., analysts are rewarded more when they are both accurate and away from the consensus), this function would have to be both path dependent (i.e., the probability of obtaining a large positive payoff increases in the number of good predictions) and asymmetric (i.e., the payoff is increasing in the number of good predictions faster than the penalty). Put differently, the payoff function should become more convex with the number of successful predictions. An example would be a system in which analysts would increase their chances of promotion by making a series of bold out-of-consensus forecasts. Although this strategy induces higher error on average, it would yield higher expected compensation because analysts would be promoted in the relatively rare cases when their forecasts are accurate (but not demoted when they are wrong).

To investigate this possibility, we consider various observable characteristics for promotion. In particular, we examine the change in the number of firms followed by the analyst, the change in the market capitalization of those firms, and whether the analyst moved to a larger brokerage house (defined as a brokerage house that follows more stocks) in the subsequent quarter. We find that none of these variables significantly affects the results. In addition, we expect our results to persist when our sample varies cross-sectionally and over time if they are explained by cognitive biases, but to change if they are driven by a particular compensation structure. The following empirical findings are more consistent with the first explanation than with the second. First, we examine whether the effect is similar for both small and large brokerage houses. This test is motivated by the understanding that compensation structures are different. To do so, we interact *STREAK* and *FREQ* with a dummy variable that equals 1 if the brokerage house employs fewer analysts than the median number. The interaction is not significant, suggesting that the effect is similar for both smaller and larger analyst firms. Second, we interact *FREQ* (or *STREAK*) with a dummy variable that takes the value of 1 if the observation is post-1992, zero otherwise. We make this time distinction because major television or Internet financial news channels did not exist in the 1980s. The interaction is not significant. This result suggests that reputation in the media is not driving our results.

6. Conclusion

Motivated by two major behavioral principles, we investigate the short-term dynamics behind analysts' forecasts and whether analysts become overconfident in their ability to predict future earnings after a series of good predictions. Overconfidence in this setting implies that analysts overweight their own estimates and rely less on public signals. Therefore, they are more likely to be out of the consensus and to have a larger prediction error in their subsequent forecast.

To test these hypotheses, we regress both departures from the forecast consensus and forecast errors on the number of "good" predictions (i.e., an error lower than the median error) in the last four quarters and with control variables. We find that both deviations from the consensus and prediction errors are positively correlated with past success. Additional control variables such as experience do not affect this bias. We explain these results by identifying overconfidence as a short-term phenomenon that recurrently appears and disappears. Its intensity varies with the length of success.

These results, however, cannot totally reject rationality because they can always be explained by making ad hoc assumptions on the reward function of the brokerage house or of the analysts. Yet, our two tests limit compensation structures that are consistent with rationality: assuming that the analysts' reward function is increasing in accuracy and that deviation from the consensus is rewarded if it is associated with lower errors, our results indicate that this function would have to be both path-dependent and asymmetric. In addition, subsequent tests provide no support that such a compensation structure explains our results: the phenomenon is persistent over time, is common to small and large firms despite differences in incentive structures and is not associated with observable proxies for promotion.

Acknowledgments
The authors would like to thank workshop participants at the University of Chicago, the 7th Annual Tel Aviv University Accounting and Finance Conference and International Association of Accounting Education and Research (IAAER) 2000 as well as Ray Ball, Daniel Bens, Robert Bloomfield, Brian Bushee, Jim Frederickson, Clive Lennox, Steven Levitt, Katherine Schipper, Richard Thaler, Richard Willis, George Wu, Robert Wyer, and numerous colleagues (especially Christopher Malloy and Cade Massey) in the Ph.D. program for their insightful comments. We are also grateful to the IAAER for awarding the Aoki Outstanding Papers Prize to this paper.

References

Abarbanell, J. S., R. Lehavy. 2003. Biased forecasts or biased earnings? The role of earnings management in explaining apparent optimism and inefficiency in analysts' earnings forecasts. *J. Accounting Econom.* **36**(1–3) 105–146.

Barber, B., T. Odean. 2001. Boys will be boys: Gender, overconfidence, and common stock investment. *Quart. J. Econom.* **1** 261–292.

Barberis, N., M. Huang. 2001. Mental accounting, loss aversion, and individual stock returns. *J. Finance* **56** 1247–1292.

Beckman, L. 1970. Effect on student's performance on teachers' and observers' attribution of causality. *J. Ed. Psych.* **61** 76–82.

Bloomfield, R. J., R. Libby, M. W. Nelson. 2000. Underreactions, overreactions and moderated confidence. *J. Financial Markets* **3** 113–137.

Daniel, K., D. Hirshleifer, A. Subrahmanyam. 1998. Investor psychology and security market under- and over-reactions. *J. Finance* **53**(5) 1839–1886.

DeBondt, W. F. M., R. M. Thaler. 1990. Do security analysts overreact? *Amer. Econom. Rev.* **80** 52–57.

Easterwood, J. C., S. R. Nutt. 1999. Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism. *J. Finance* 1777–1797.

Erev, I., T. S. Wallsten, D. V. Budescu. 1994. Simultaneous over- and underconfidence: The role of error in judgment processes. *Psych. Rev.* **101** 519–528.

Fama, E. F., K. R. French. 2000. Forecasting profitability and earnings. *J. Bus.* **73**(2) 161–175.

Fama, E. F., J. D. MacBeth. 1973. Risk, return and equilibrium: Empirical tests. *J. Political Econom.* **81** 607–636.

Fiske, S., S. E. Taylor. 1991. *Social Cognition*. McGraw Hill, New York.

Fitch, G. 1970. Effect of self-esteem, perceived performance, choice on causal attributions. *J. Personality Soc. Psych.* **16** 311–315.

Francis, J., D. Philbrick. 1993. Analysts' decisions as products of a multi-task environment. *J. Accounting Res.* **31** 216–230.

Gervais, S., T. Odean. 2001. Learning to be overconfident. *Rev. Financial Stud.* **14**(1) 1–27.

Gu, Z., J. S. Wu. 2003. Earnings skewness and analyst forecast bias. *J. Accounting Econom.* **35**(1) 1–116.

Harvey, N., C. Harries, I. Fischer. 2000. Using advice and assessing its quality. *Organ. Behav. Human Decision Processes* **81** 252–273.

Hastorf, A., D. Schneider, J. Polefka. 1970. *Person Perception*. Addison-Wesley, Reading, PA.

Hong, H., J. D. Kubik, A. Solomon. 2000. Security analysts' career concerns and herding of earnings forecasts. *RAND J. Econom.* **31**(1) 121–144.

Huck, S., J. Oechssler. 2000. Informational cascade in the laboratory: Do they occur for the right reasons. *J. Econom. Psych.* **21** 661–671.

Hung, A., C. Plott. 2001. Information cascades: Replication and an extension to majority rule and conformity rewarding institutions. *Amer. Econom. Rev.* **91**(5) 1508–1520.

Jacob, J., T. Z. Lys, M. A. Neale. 1999. Expertise in forecasting performance of security analysts. *J. Accounting Econom.* **28** 51–82.

Johnson, T. J., R. Feigenbaum, M. Weiby. 1964. Some determinants and consequences of the teachers' perception of causation. *J. Ed. Psych.* **55** 237–246.

Klayman, J., J. B. Soll, C. Gonzalez-Vallejo, S. Barlas. 1999. Overconfidence: It depends on how, what, and whom you ask. *Organ. Behav. Human Decision Processes* **79**(3) 216–247.

Koehler, D. J., L. Brenner, D. Griffin. 2002. The calibration of expert judgment: Heuristics and biases beyond the laboratory. T. Gilovich, D. Griffin, D. Kahneman, eds. *Heuristics and Biases*. Cambridge University Press, Cambridge, UK.

Kraemer, C., M. Noeth, M. Weber. 2006. Information aggregation with costly information and random ordering: Experimental evidence. *J. Econom. Behav. Organ.* **59**(3) 423–432.

Kukla, A. 1972. Attribution determinants of achievement-related behavior. *J. Personality Soc. Psych.* **21** 197–207.

Kunda, Z. 1990. The case for motivated reasoning. *Psych. Bull.* **108**(3) 480–498.

Larrick, R. P., J. B. Soll. 2004. Strategies for revising judgement. Working paper, Duke University, Durham, NC.

Mikhail, M. B., B. R. Walther, R. H. Willis. 1997. The development of expertise: Do security analysts improve their performance with experience? *J. Accounting Res.* **3** 131–157.

Miller, D. T. 1976. Ego involvement and attributions for success and failures. *J. Personality Soc. Psych.* **434** 901–906.

Phillips, D. J., E. W. Zuckerman. 2001. Middle-status conformity: Theoretical restatement and empirical demonstration in two markets. *Amer. J. Sociology* **107**(2) 379–429.

Richardson, S., S. Hong Teoh, P. Wysocki. 2004. The walkdown to beatable analyst forecasts: The role of equity issuance and insider trading incentives. *Contemporary Accounting Res.* **21**(4) 885–924.

Sinha, P., L. D. Brown, S. Das. 1997. A re-examination of financial analysts' differential earnings forecasts accuracy. *Contemporary Accounting Res.* **14**(1) 1–42.

Weiner, B., A. Kukla. 1970. An attributional analysis of achievement motivation. *J. Personality Soc. Psych.* **15** 1–20.

Yaniv, I. 2004. Receiving other people's advice: Influence and benefit. *Organ. Behav. Human Decision Processes* **94**(1) 1–13.