Discussion of "The Effects of Financial Statement Information Proximity and Feedback on Cash Flow Forecasts"*

JOAN LUFT, *Michigan State University*

1. Introduction

In this study Hodge, Hopkins, and Wood (hereafter HHW) provide experimental evidence related to two basic financial statement format issues: the location of performance information and the number of periods of performance information included. While HHW focus on their experiment's contribution to current debates in standard setting, the primary intent of this discussion is to supplement their presentation by considering more general implications of their study. (I will, however, return to the standard-setting issues at the end.)

The questions raised in HHW are relevant to the general problem of learning from accounting information in a changing environment. Which presentation formats will better enable people to see significant patterns in accounting information? Which presentation formats will enable them to see these patterns more *quickly*? In a competitive environment, it is not enough that learning be accurate: it must also be fast.

HHW make two interesting and (to the best of my knowledge) novel points about financial statement presentation. First, they remind us that reading financial statements is not only a cognitive process, it is also a *physical* process. The location of items in financial statements influences judgments, not only because location has informational implications (e.g., as an indicator of reliability or similarity to neighboring items), but also because location constrains the physical process of looking, which in turn affects judgment. Information that can be "taken in at a glance" is likely to be integrated more effectively in judgments than information that is physically far enough apart to require multiple glances.

The second novel point in the paper is that *repetition* of feedback could influence judgment. A primary difference between the limited and extensive feedback condition in HHW is that the limited feedback condition presents participants with feedback about any given period only once, just after the relevant forecast has been made, while the extensive feedback condition presents feedback for each period three times, each time in conjunction with feedback from two other periods.

* Accepted by Hun-Tong Tan. I am grateful to the editor and to participants at the CAR conference for their comments.

Contemporary Accounting Research Vol. 27 No. 1 (Spring 2010) pp. 135–142 © CAAA
doi:10.1111/j.1911-3846.2010.01004.x

These are potentially important issues, but skeptical readers may wonder how much an experiment like HHW, which uses student participants, modest incentives, and a highly stylized, simplified task, and yields somewhat mixed results, can tell us about the real-world impact of these differences in financial statement presentation. This discussion is therefore organized around three questions that are likely to be the focus of skepticism about the study.

(a) Experiment participants are paid for performance, but their pay is very small compared to the sums at stake in real-world investment choices. Would the effects documented in this study be likely to disappear if the incentives for accurate forecasting were higher, and if so, are the experimental results irrelevant?
(b) By the last of the sixteen trials, median forecast error is similar, and relatively low, in all experimental conditions. Would we not expect that forecasts by more experienced nonprofessional investors in the real world — who have probably looked at more than sixteen instances of cash flow in their lifetimes — would be much like trial-sixteen forecasts: that is, they would be unaffected by proximity or feedback differences?
(c) The feedback manipulation does not have most of the predicted effects on judgment (no main effect on forecast error or variability, no effect on the speed of reduction of forecast error, and only a marginally significant effect on the speed of reduction of variability). However, there are (non-hypothesized) three-way interactions of proximity, feedback, and trial, affecting variability and time spent on the task. Does the feedback manipulation actually have any intelligible effect? Each of these three questions is addressed in more detail below.

2. Are the incentives too small?

The human factors research that HHW cite in their hypothesis motivation has documented the importance of apparently small information-location differences in high-stakes settings like piloting aircraft, performing heart surgery, and managing nuclear plants. It might be argued that if low proximity is a real problem in life-and-death settings, where incentives are unquestionably high, then it is likely also to be a real problem in financial settings. The weakness of this argument, however, is that in the dramatic settings of human factors research, decisions must sometimes be made extremely quickly, and the few seconds needed to scan for physically distant information can therefore be very costly. HHW's task is less urgent. Nonprofessional investors often have time to look back and forth at different statements, to put a copy of one statement side by side with another, or even reformat electronically available information in order to bring related information into proximity. Are they not likely to do so if it is worthwhile to them?

The extreme simplicity of HHW's experiment helps to address this question. To see how it does so, recall the experimental task: participants forecast cash flow from operations (CFO) at time $t + 1$ based on only two pieces of information, CFO at time t and change in net operating assets (CNOA) at time t. The two pieces of information appear either on the same screen or on succeeding screens. Participants learn accurate forecasting very quickly when they can see both predictor variables on the same screen. If participants in the low proximity condition feel disadvantaged by the visual separation of the two variables, then they could note the first number on a scrap of paper and hold it up next to the second number when the second number appears on the screen. If this practice is discouraged by the experimenter or seems inappropriate to them, then they can memorize the first number and project a mental picture of it next to the second number. (This may be less satisfactory than the first alternative, but it is probably better than doing nothing to reformat the data presentation.)

Evidently, since participants' judgments are significantly less accurate in the low-proximity condition, they are not taking successful action of this kind. How much money are they leaving on the table by not making this effort? That is, how small are the real incentives for making this effort? In trying the experimental task, I found that noting or effectively memorizing CFO_t added three to four seconds to the time required per trial. Thus the strategy required to mitigate the effect of low proximity would add about a minute (4 seconds x 16 trials = 64 seconds) to task time. Participants' pay is based on forecast error, and moving from the low proximity to high proximity conditions cuts forecast error approximately in half (Table 2 in HHW). It is difficult to be sure exactly how much this error reduction adds to participants' pay; but we might reasonably suppose, given the actual pay range of \$6 to \$20, that cutting error in half might add an expected \$3–\$4 to pay. A rate of \$3-\$4 for a minute of additional effort translates to \$180–\$240 per hour. It could be argued that this is not a trivial incentive.

The obstacle to accurate forecasts in the low proximity condition is probably not participants' unwillingness to take the effort to reformat the data, however. It is their lack of awareness that doing so would significantly improve their forecasts. When we think about cognitive costs, it is important to remember that these include both the cost of *using a decision strategy*, which can be very low, and the cost of *knowing which decision strategy* to use, which can be much higher.

To pursue this argument further, recall that by the time participants finish HHW's experiment they forecast cash flows relatively accurately even in the low proximity condition. Suppose, therefore, that they have acquired effective judgment strategies by this time. (This is a strong assumption, which section 3 will question; but for purposes of the present argument, it provides a conservative estimate of the "cost of knowing.") If the cost of knowing is \$180–\$240 per hour — the amount of money low proximity participants have left on the table in this experiment but would not leave on

the table in the future — and it can be amortized over many future decisions, then it is very small on a per-decision basis. Is this cost therefore trivial, compared to the benefits of better decisions?

A reasonable answer to this question needs to take into account the facts that (a) acquiring and examining cash flow information for sixteen real firms will take nonprofessional investors considerably longer than one minute, and (b) more than sixteen instances would be required to learn to forecast accurately in the more uncertain real-world setting (see section 3 for further comment on this point). Suppose, for the sake of argument, that the cost of time, effort, and possibly advice required for investors with low proximity information to learn to forecast as well as those with high proximity information is $200. Suppose further that the lifetime payoff to this effort is much higher than $200, and therefore investors with low proximity information typically *do* reformat it and therefore make the same judgments and earn the same returns as they would with high proximity information. Should we then conclude that proximity is not important, because investors know how to overcome its effects and the incentives are sufficient for them to do so? Perhaps not.

If low proximity imposes the somewhat trivial one-time processing cost of $200 on each of over 40 million nonprofessional investors, and the cost to preparers of placing information in one location (high proximity) is no greater than the cost of placing it in another location (low proximity), then the total effect of standard setters' choosing low proximity would be a dead-weight loss of over $8 billion. It is unlikely, of course, that these are accurate estimates of the costs of either processing low-proximity information or providing high-proximity information. But the point remains: even if incentives in the environment are sufficient to motivate investors to incur additional cognitive costs and overcome judgment biases demonstrated in the laboratory, this does not necessarily mean that the costs identified in the laboratory are irrelevant.

3. Would more experienced users be unaffected by format?

The experiment participants are young, with little investment experience. They make large forecast errors in the first trials of the experiment, but by trial 16 median forecast error is low, and probably not significantly different, across all experimental conditions. Many nonprofessional investors in the real world have seen more than sixteen instances of cash flow from operations in their lives. Should we therefore conclude that proximity will be unlikely to matter to their judgments?

Two lines of argument are relevant to this question. First, it is much easier for participants in this experiment to learn accurate cash-flow forecasting from experience than for real-world investors to do so, because the data used in the experiment are far less noisy than real-world data. The adjusted R^2 of a regression using the data provided to participants in HHW (Table 2) is over 98%, while comparable models in Barth, Cram and Nelson

2001, which are estimated from archival data and provide the basis for constructing the experimental materials in HHW, have adjusted R^2s below 35%. Learning takes longer when environmental predictability is lower, and the predictability implied by the Barth et al. 2001 models is low enough to challenge the limits of individuals' ability to learn from feedback (see Hogarth and Karelaia 2008 for a review of the relevant literature).

Second, it appears that learning is not a very durable good in this setting. Even when individuals *have* learned to forecast more accurately, and when median forecast error appears to have stabilized at a low level, proximity-based judgment problems are not at an end. This is evident from recurring error spikes in later periods in the low proximity cells after sequences of stable low-error trials (trials 6–11 in low proximity limited feedback and 7–14 in low proximity extensive feedback; see Figure 2 in HHW).

Even if participants *have* (in some intuitive, implicit sense) learned the coefficients on CFO_t and $CNOA_t$ and if this knowledge is stable by trial 16, it does not necessarily follow that participants have learned how to learn coefficients in general, or that they have learned that low proximity is a judgment hazard. They may have no more idea than they did at the beginning of the experiment that high proximity improves their forecasts, and that they would make more money in a low proximity setting in the future if they mentally projected CFO_t next to $CNOA_t$ on the screen when making their forecast.

This lack of higher-order learning is important because the coefficients themselves also are not necessarily durable goods. Historical changes in accounting practice or the economic environment could change the optimal coefficients on CFO_t and $CNOA_t$ in a predictive model. Moreover, if nonprofessionals begin by attending only to some industries and later expand or otherwise change the scope of their interests, they will have to learn new coefficients (see Barth et al. 2001, Table 6, for examples of the large cross-industry variation in these coefficients). Experienced nonprofessional investors will learn new coefficients as badly as HHW's low proximity participants if their experience has taught them only that "the coefficient on CFO_t is 1", not that "predictors should be brought into proximity".

4. What effects does the feedback manipulation have?

An important role of accounting is the provision of outcome feedback from which investors, managers, and others can learn the effects of investment and operating decisions and the patterns of relation over time among multiple financial measures such as cash flows and accruals. Given the importance of this role, it is reasonable to ask whether different ways of presenting such accounting feedback make a significant difference to learning and judgment performance.

HHW's feedback manipulation does more to raise questions about feedback effects than it does to provide answers. A number of the hypothesized

effects do not occur: feedback type has no significant effect on median forecast error or variability, nor on the speed of reduction in forecast error, although there is a marginal effect on the speed of reduction of variability. Three-way interactions of proximity, feedback, and trials limit the interpretability of lower-order effects (main effects and two-way interactions) in the analyses of variability and time spent on the financial statements.

The three-way interactions are not theorized, and it is not evident (at least to me) why they take the forms they do. It appears from Figures 3 and 4 that more extensive feedback increases the speed of learning to judge *similarly* (lowering variability) in high proximity but has little effect in low proximity. In contrast, more extensive feedback appears to increase the speed of learning to judge *quickly* (time spent on the financial statements) in low proximity but not in high proximity.

One of the difficulties with theorizing the effects of the feedback manipulation is that it has several components with potentially different effects. The difference between limited and extensive feedback includes the following characteristics: in extensive feedback (a) three periods are presented instead of one, (b) feedback for any given period is *repeated*: feedback for a period is presented three times instead of once, and with each of the multiple repetitions it is presented with a different set of other periods, and (c) predictor variables are presented along with predicted and actual CFO, while limited feedback presents predicted and actual CFO only. Insofar as there *are* results of the feedback manipulation, we do not know which of these features is driving the results.

This rather complicated manipulation has a benefit, however, in confronting us with potentially important questions about outcome feedback presentation. Accounting research has tended to mimic the highly stylized presentation of outcome feedback in psychology research and has therefore done relatively little to investigate the multidimensional variation in how accounting feedback is actually presented. In multiple cue probability learning studies in psychology, individuals typically make a prediction and then see the actual outcome in each trial, sometimes with no opportunity to look back at earlier individual trials or examine several trials at once. This standard experimental format (which resembles but is not identical to HHW's limited feedback) is perhaps not very conducive to learning and is certainly not representative of many naturally occurring learning opportunities in the accounting domain.[1] Investors often make forecasts for several firms at once and thus receive outcome feedback for several predictions at roughly the same time, facilitating comparisons across trials. For investors who

1. Luft and Shields (2001), investigating a management accounting setting in which individuals learn the relation between prior expenditures and profits from information received simultaneously from a number of similar business units, provide a brief discussion of potential differences between their setting and the traditional limited-feedback setting of most multiple cue probability learning studies.

follow a firm over time, the presentation of multiple instances of past predictors and actual values in financial statements also provides repetitions of previously encountered instances, perhaps facilitating a regular reconsideration of previous judgments that is unlikely to occur with conventional psychology experiment feedback.

HHW thus make a general contribution by prompting accounting researchers to think about the actual properties of accounting feedback and how these properties might help or hinder users' learning. The specific contribution of the feedback manipulation to the standard-setting debate is more questionable, however. Insofar as extensive feedback does support better learning (e.g., faster reduction of variability in high proximity), we cannot be sure whether this is due to properties that the manipulation shares with real-world accounting feedback or to properties it does not share.

Two properties of the feedback in HHW differentiate it from the feedback provided by financial statements. First, the standard-setting issue to which HHW refer is a question about how many periods of performance information to provide. But the limited feedback in HHW differs from the extensive feedback not only with respect to the number of periods but also with respect to the amount of information provided per period. Extensive feedback provides the previous predictors (CFO_t and $CNOA_t$), as well as the actual CFO_{t+1} and the participant's forecast and percentage error, thus allowing participants to easily reconstruct and evaluate their previous judgments. In contrast, limited feedback does not provide the previous predictors, but only the forecast, actual, and percentage error. With HHW's limited feedback participants can see *that* they have made a good or bad forecast, but without the prompts provided by the previous predictors, they may be less able to recall *how* they made the forecast and thus less able to correct their forecasting strategy. If this matters — and we do not know whether it does or not: this is an open and perhaps interesting question — then such feedback results as we do see in HHW are not unambiguously informative about the effect of number of periods as such.

Second, unlike both the limited and extensive feedback in HHW, accounting reports do not include the user's own previous forecasts and percentage forecast errors (cognitive feedback, in addition to the outcome feedback provided by actual CFO_{t+1}). These salient pointers toward possible weaknesses in the individual's judgment strategies might provide significant supports for learning that are not provided by actual financial statements. In the absence of such supports, by the time real-world nonprofessional investors see CFO_{t+1} they may remember their original forecasts imperfectly and (given individuals' tendency to hindsight bias) they may believe that these forecasts were closer to the actual outcome than they really were. If they overestimate the success of their forecasting strategy in this way, they will not correct it as quickly as HHW's participants do.

It should be noted, however, that even if cognitive feedback significantly influences individuals' learning, this does not necessarily mean that HHW's proximity results are uninformative about the effects of proximity in actual financial statements that lack this cognitive feedback. A threat to generalizability exists only if the presence of cognitive feedback *interacts with* proximity. If cognitive feedback has only a main effect — that is, if absence of cognitive feedback degrades judgment performance in all conditions by a similar amount but leaves the difference between high and low proximity unchanged — then HHW's results remain informative even for nonprofessional investors who do not receive cognitive feedback and do not create it for themselves by retrieving previous forecasts and calculating their percentage errors. Whether cognitive feedback actually *does* interact with format effects like proximity is another of the many things we do not know about how individuals learn from accounting feedback.

5. Conclusion

In sum, what does this study contribute? First, HHW remind us that judgments and decisions are physical processes — not only the processes of neurons firing in the brain, but also the processes of looking and page-turning, pointing and clicking — and thus the physical properties of financial statements (whether paper or electronic) are relevant as well as their informational properties. Second, this study reminds us that learning is important, multidimensional (learning a coefficient is different from learning the importance of proximity), and imperfectly understood. We see that learning from accounting feedback is significant but limited, and it is influenced by accounting presentation in ways that can be clarified only through further research.

References

Barth, M. E., D. P. Cram, and K. K. Nelson. 2001. Accruals and the prediction of future cash flows. *Accounting Review* 76 (1): 27–58.

Hodge, F. D., P. E. Hopkins, and D. A. Wood. 2010. The effect of financial statement information proximity and feedback on cash flow forecasts. *Contemporary Accounting Research* 27 (1): 101–33.

Hogarth, R. M., and N. Karelaia. 2008. Determinants of linear judgment: A meta-analysis of lens-model studies. *Psychological Bulletin* 134 (3): 404–26.

Luft, J., and M. D. Shields. 2001. Why does fixation persist? Experimental evidence on the judgment performance effects of expensing intangibles. *Accounting Review* 76 (4): 561–87.