The performance effect of feedback frequency and detail: Evidence from a field experiment in customer satisfaction[1]

March 2016

Pablo Casas-Arce
Arizona State University
casas.arce@gmail.com

Sofia M. Lourenço
ISEG, Universidade de Lisboa
and Advance, CSG Research Center
slourenco@iseg.ulisboa.pt

F. Asís Martínez-Jerez
University of Notre Dame
asismartinez@nd.edu

Abstract: This paper presents the results from a field experiment that examines the effects of non-financial performance feedback on the behavior of professionals working for an insurance repair company. We vary the frequency (weekly and monthly) and the level of detail of the feedback that the 800 professionals receive. Contrary to what we would expect if these professionals conformed to the model of the Bayesian decision maker, more (and more frequent) information does not always help improve performance. In fact, we find that professionals achieve the best outcomes when they receive detailed but infrequent (monthly) feedback. The treatment groups with frequent feedback, regardless of how detailed it is, perform no better than the control group (with monthly and aggregate information). The results are consistent with the information in the latest feedback report being most salient, and professionals in the weekly treatments overweighting their most recent performance, hampering their ability to learn.

[1] We thank Larbi Alaoui, Margaret Christ, Juanjo Ganuza, Gary Hecht, Chris Ittner, Michal Matejka, Andrei Shleifer, and seminar participants at the AAA Annual Meeting, IAP Cambridge Research Symposium, IE, MAS Midyear Meeting, Northwestern University, Temple University Accounting Conference, Universidad Complutense de Madrid, Universitat Pompeu Fabra, University of Notre Dame, and University of Toledo for helpful comments. Sofia M. Lourenço gratefully acknowledges financial support from FCT—*Fundação para a Ciência e Tecnologia* (Portugal), research grant PTDC/EGE-GES/119607/2010 (national funding). All errors are our own.

I. Introduction

One of the main roles of accounting information is to facilitate decision-making. Timeliness and detail are usually regarded as desirable characteristics of information because they enable prompt and adequate responses to business threats and opportunities. However, the intensity of these attributes must be weighed against the decision maker's ability to process the relevant information. Too frequent information may result, for instance, in an overreaction to short-term factors, whereas too detailed information may cloud a decision maker's ability to identify general trends or issues. In this paper, we use a natural field experiment to analyze how the frequency and detail of performance feedback influences employee behavior.

Natural field experiments like ours, although rare in the accounting literature, have the advantage of combining the most attractive elements of experimental and archival research: control and realism (Floyd and List 2016). The random allocation of subjects to a feedback treatment allows us to attribute causality to the stimuli without the identification challenges posed by self-selection and the presence of confounding factors in naturally occurring data. Furthermore, conducting the experiment in the field provides a realistic institutional context that is hard to replicate in the lab. Our subjects are professionals in their field going about their daily jobs. They have implicit and explicit incentives linked to their performance. Moreover, they were not aware that they were experimental participants and, thus, their behavior is not contaminated by speculation about the experimenters' objectives. Finally, the theoretical foundation of our experiment design allows us to generalize the insights that we glean from our results.

In collaboration with Multiasistencia—the leading Spanish business process outsourcer of repairs for insurance companies—we design and implement a field experiment in which we manipulate the non-financial performance feedback received by 800 home repair professionals (such as plumbers, masons, or painters) who work with the firm. A natural field experiment is especially well suited for this context. Our subjects are professional decision makers who use this information in their daily decisions and understand how that information can be used to fine-tune their actions and behaviors. The richness of their customer interactions exceeds what can be codified in objective indicators and emphasizes the importance of the local, "soft" information that the professionals capture in their daily activities. This type of information cannot be replicated in a lab environment, but it is instrumental in the decision makers' interpretation of satisfaction scores.

We actively intervene in the reward and feedback system by introducing a bonus that rewards the achievement of a target for customer satisfaction—a metric that has been found to be a significant leading indicator of the financial health of a firm (Ittner and Larcker 1998)—as well as two process indicators. We vary the frequency of feedback information (weekly vs. monthly) and the level of detail included in the report (the average score of all jobs performed during the period by a professional vs. the scores for each of the individual jobs).

If the professionals conformed to the ideal of the Bayesian decision maker, they would efficiently use all the information available to them to improve customer satisfaction. Therefore, more detailed and more frequent feedback should lead to better performance. However, more (and more frequent) customer satisfaction feedback does *not* always result in improved customer satisfaction scores. In fact, we find that professionals achieve higher scores when they receive detailed but infrequent (monthly) feedback. These results are consistent with the latest feedback report being most salient, and professionals overweighting the information it contains. As a result, although detailed customer satisfaction feedback supplies information that helps professionals to improve the service they provide, feedback that is more frequent (and that consequently focuses on a shorter time horizon) ends up harming service, as previous information is disregarded in the face of new information.

Notably, we also show that the deterioration in performance for the professionals in the weekly treatments is mainly explained by an overweighting of the most recent performance information and not by the rational abandonment of the pursuit of the monthly bonus. We dismiss the presence of dynamic incentives as the main driver of our results by showing that weekly treatments do worse than monthly treatments in the first week of the month, particularly when they follow a negative report in the last week of the prior month. This result cannot be explained by the rational abandonment story, as bad performance in the last week of a month has no impact on the chances of achieving the bonus in the next month.[2]

These differences among treatments do not exist with respect to the process indicators included in the bonus system (e.g., the use of the Internet to schedule a service or finishing a repair on time). This is because the professionals receive immediate feedback simply by executing these tasks. Thus, the differences in the features of the formal feedback system do not result in any additional information, and do not affect professionals' knowledge about their performance or the way they process that information (Annett 1969). Consistent with this, we show that operational performance is indeed the same in all four treatments.

This paper contributes to several streams of literature. First, we contribute to the literature in information economics. In this literature, the contracting stream of research on performance measurement mainly focuses on how properties of information affect their inclusion in contracts (Feltham and Xie 1994; Prendergast 2002; Moers 2006). The ultimate objective of this literature is to judge the strength of the performance metric in providing information to the *firm* about employees' choices (the control function). The design of performance metrics to facilitate the

---

[2] Moreover, there is no reason for professionals in the weekly treatments to be more likely to infer that the bonus is not achievable. Both weekly and monthly treatments have the same information about past performance during the first week of each month.

employee's decision-making (the decision-making function) has been analyzed in the literature only rarely (Sprinkle 2003; Casas-Arce et al. 2016). In this study, we keep constant the incentive compensation and show how changes in the detail and frequency of performance metrics affect decision makers' behavior.

Our work also contributes to the feedback literature. Although traditionally more feedback is viewed as having a positive effect on performance, Kluger and DeNisi (1996) argue that the evidence is not always consistent with this view, and that a lack of theory has hampered our understanding of the factors that condition its effectiveness. Our paper provides further evidence that more feedback information does not necessarily lead to better performance. Furthermore, we develop a theory based on information salience that is amenable to studying the effects of feedback, and that is able to accurately explain our evidence.

Although economists often assume that people make rational inferences from all available information (Savage 1954), psychologists have provided ample experimental evidence that is inconsistent with rational decision-making in general, and Bayesian inference in particular (see Kahneman and Tversky 1972, 1974, 1983). For instance, individuals tend to overweight the data that is most salient in order to economize limited cognitive resources (Taylor and Thompson 1982). Although many other biases in decision-making have been uncovered, recent work by Gennaioli and Shleifer (2010) and Bordalo, Gennaioli, and Shleifer (2012, 2013) shows that salience can account for a number of these behavioral anomalies and explain behavior in a wide range of settings. Our evidence shows that feedback in organizations can similarly induce behavioral responses that are not consistent with a Bayesian decision maker, but that can easily be accounted for by assuming that feedback reports are salient. In particular, infrequent feedback increases the professional's ability to process information (especially if the feedback contains detailed information) and improves his or her decision-making. Our results speak to the importance of the design of accounting information systems in organizations.

The structure of the paper is as follows. Section II reviews the relevant literature. Section III provides the institutional background of the research site, Multiasistencia, and the market in which it operates. The design of the field experiment is described in Section IV. Section V provides motivation for the empirical tests. Section VI analyzes the empirical results, and Section VIII concludes.

II. Review of Feedback Research

The traditional view in the literature is that feedback leads to performance improvement. In economic models of Bayesian updating, learning is a by-product of the utility maximization process in which a rational agent uses the new information provided by feedback to update her beliefs about the probable consequences of her choices and the impact on her utility (Savage 1954; Kiefer and Nyarko 1995). In the management literature, feedback is considered to have a positive impact on performance because it improves learning and motivation (Ammons 1956; Ilgen et al. 1979; Kopelman 1986). However, a century-long body of research has shown that feedback does not uniformly improve performance (Balcazar et al. 1985; Kluger and DeNisi 1996; Alvero et al. 2001). There is now a consensus that the effect of feedback is contingent on the organizational setting in which it is provided and on the characteristics of the feedback itself (Balcazar et al. 1985; Kluger and DeNisi 1996). In particular, goal-setting and incentives stand out as features that appear to increase recipients' attention to feedback and improve the consistency of its effects (Locke and Latham 1990; Kluger and DeNisi 1996; Sprinkle 2000).

The specific feedback characteristics that researchers have looked at include, for example, the credibility and power of the source (Ilgen et al. 1979), whether the feedback is on individual or relative performance (Hannan et al. 2008), whether it is communicated privately to the recipient or made public (Hannan et al. 2013; Tafkov 2013), and whether it conveys a positive or negative message (Illies and Judge 2005). Two characteristics that have received special attention, and are particularly interesting for uncovering the effect of salience, are the detail and frequency of feedback. The literature has long presumed that, in line with the Bayesian updating view, more detailed and more frequent feedback improves performance. However, there are behavioral reasons why the excess of these characteristics may hamper the recipient's ability to process feedback information.

The traditional view of feedback detail is that an increase in detail improves performance. Thorndike's law of effect (1927) suggests that this is so because more detail permits a better identification of the behaviors that are reinforced and those that are punished. Detail also enhances the credibility of feedback, which becomes more believable when it is supported by specific examples (Leskcovec 1967). However, behavioral theories have questioned the positive effects of feedback detail. Very detailed feedback may direct the recipient's attention to specific events and result in the inappropriate generalization of a small number of salient situations rather than in a balanced learning inferred from all the information available, a phenomenon known as the law of small numbers (Tversky and Kahneman 1971; Rabin 2002). Moreover, when feedback provides very specific cues on how to improve performance, the recipient may disengage from the learning process, relying exclusively on the cues from feedback (Goodman et al. 2004).

Empirical evidence on the impact of feedback detail on performance is mixed: while some studies see a positive relationship, others do not, and some even find a U-shaped relationship between detail and performance (Goodman et al. 2004; Bilodeau 1969; Salmoni et al. 1984). This lack of consistency is caused in part by diversity in the definition of "detail," which can refer to traits as different as the level of precision of the feedback itself (Hannan et al. 2008) or the inclusion of advice on how to improve performance (Kim 1984). Also contributing to the lack of consistency are the different choices for the organizational design elements that interact with feedback, such as the incentive scheme (Northcraft et al. 2011; Hannan et al. 2008).

As in the case of feedback detail, the traditional view of feedback frequency in the literature is that more is better. From a learning standpoint, more frequent feedback allows the decision maker to revise her beliefs and try new strategies more often (Salmoni et al. 1984; Schmidt and Dolis 2009). From a motivational perspective, it contributes to the recipient's development of a sense of competence by allowing her to observe that her actions influence performance (Ilgen et al. 1979). Moreover, from an organizational point of view, an implicit value is given to metrics that are measured more frequently, which keeps the organization focused on those metrics (Reichheld 2006). However, behavioral theories argue that more frequent feedback may cause the recipient to lose perspective and pay more attention to the most recent performance. This orientation encourages a fire-fighting approach to problem solving rather than a long-term fundamental approach (Bohn 2000; Lurie and Swaminathan 2008).[3] Additionally, more frequent feedback also increases the noise of the performance signal and could make it more difficult to learn (Bohn 1995; Lurie and Swaminathan 2008).[4]

[3] A parallel line of argument exists in the disclosure literature. Van Buskirk (2012) finds that more frequent disclosure leads to more speculation by investors. Bushee and Noe (2000) find that increases in a firm's disclosures—as measured by AIMR disclosure rankings—are associated with increases in speculative trading by institutional investors.

[4] Lurie and Swaminathan (2008) also find that frequent feedback may lead to poor performance. However, there are several significant differences between our studies. First, although their findings—and virtually all of the feedback literature—use a laboratory experiment, we implement our treatments in a field experiment. We observe

Although some experiments suggest that more frequent feedback may not improve performance (Chhokar and Wallin 1984, Lurie and Swaminathan 2008), most of the studies support the positive performance effects of frequent feedback (Kluger and DeNisi 1996; Balcazar et al. 1985; Alvero et al. 2001; Northcraft et al. 2011; Kang et al. 2005). A common explanation for the inconsistent results of these studies is that they suffer from methodological problems because they do not test purely for frequency but also add level of detail and/or other reinforcers such as training in the treatments.

Previous studies have mainly looked at feedback frequency and detail independently (e.g., Goodman et al. 2004; Chhokar and Wallin 1984). Northcraft, Schmidt and Ashford (2011) are an exception in the sense that they look at the joint effect of both characteristics, but their lab experiment does not focus on how these characteristics affect the processing of information. Rather, they examine how the combination of feedback frequency and detail affects the salience of competing tasks and how decision makers allocate resources among those tasks.[5]

III. Research Setting

---

professionals operating in a real, high-stakes setting. This setting is especially relevant for accounting research, as it involves incentives linked to the achievement of performance targets. The incentive structure in our setting allows us to distinguish between rational abandonment and the salience hypothesis. Second, we show that the evidence is inconsistent with a rational explanation and provide an alternative theoretical framework that tightly accounts for our empirical results. Finally, the nature of the task and the design of the reporting system treatments in our study are well suited for testing the salience hypothesis. In our setting, professionals only receive outcome feedback for the last reporting period. Decision makers thus have to exercise judgment in making inferences from that information to improve performance. In contrast, Lurie and Swaminathan's (2008) subjects receive feedback on performance and on the process that led to that performance. Their feedback reports also include information on all past actions, reducing the salience of recent performance.

[5] Their lab experiment asks subjects to perform four simple tasks simultaneously, with each task receiving a different feedback treatment. They study how the feedback characteristics affect the decision makers' allocation of attention among those tasks. In contrast, we use a field experiment to examine how detail and frequency affect decision makers' ability to process the relevant information to learn how to improve performance in the execution of a single job.

Multiasistencia is a business process outsourcing (BPO) firm that provides comprehensive claims management service for property and casualty insurance companies. The firm acts as the coordinator between clients with repair needs and a network of specialized home repair professionals. It is located in Europe and Latin America and is the industry leader in Spain, the country in which we base our study.

Multiasistencia's largest corporate clients in Spain are the insurance subsidiaries of major banks. The insurance companies hire Multiasistencia to manage the claims process for individual properties from the first report by the customer to the finishing touches of the repair.[6] Client relationships are governed by annual contracts. Typically, performance is formally reviewed on a monthly basis against service level agreements (SLAs) that include parameters of cost, timeliness, and quality of service. The CEO explained the nature of the interaction thus: "We assign a key account manager to each of the major insurance companies to oversee that client's specific needs. The management team also maintains close connections with our largest corporate clients and communicates with their leaders approximately once a week."

Multiasistencia employs over 300 customer service representatives (CSRs) in its call centers. There are separate phone banks for each of the four largest corporate clients and one general phone bank for overflow calls and calls from smaller customers. In a typical service intervention, the policyholder reports a claim by calling the insurance company, which redirects the call to Multiasistencia. The CSR at the call center makes an initial assessment of whether the caller's claim is covered by the policy that he or she holds. Claims deemed to be covered by the policy are transferred to a regional dispatch office, where jobs are assigned to repair professionals as a

[6] We use the term "client" to refer to the insurance companies that outsource their repair work to Multiasistencia and the term "customer" to refer to the policyholder.

function of the expertise required for the repair and the workload of the professional. Information from each call is recorded in a computer database. From a policyholder's point of view, the entire process is managed through the insurance company that has delegated its repair work to Multiasistencia.

Small repairs (less than three man-hours) are assigned to a repair professional who confirms or denies the coverage of the reported damage. If the professional confirms the claim is covered by the policy, he or she carries out the repairs, completes a report, and closes the job in one visit. For larger jobs, the CSR assigns a professional to repair urgent damages and orders an assessment for the rest of the job. A claims inspector is sent to the site within two days of the call and issues a report to Multiasistencia. If the report justifies the claim, the professional is sent to complete the rest of the repairs. For repairs requiring more than one specialty (e.g., plumbing and glass repair), the intervention of each professional is scheduled sequentially by the dispatch center. Workflow and communications with and among professionals are managed and recorded through a system of hand-held devices (PDAs) supplied by Multiasistencia. At the end of each repair, a CSR contacts the policyholder to check that the repair has been completed.

The Repair Professionals

Multiasistencia works with a network of professionals. Repair professionals are not direct employees of Multiasistencia but are linked to the firm by relational contracts through which they receive a guaranteed stream of jobs. In exchange for receiving a guaranteed workflow, repair professionals commit to following Multiasistencia's operational procedures and prioritizing the firm's repairs.

Professionals are paid a fixed fee for visits that result in denial of coverage and for small jobs. For large jobs (those involving more than three hours of work), they are compensated on a variable scale based on the cost of materials and the number of hours needed to complete the repair. Small jobs account for 80% of all approved claims.

Prior to our experiment, there was no explicit incentive compensation system in place for repair professionals. However, Multiasistencia did track a set of operating indicators at the professional level.[7] Regional managers told professionals which indicators needed more of their attention, and better performers were implicitly rewarded with a heavier stream of work.

Customer Satisfaction

In 2012, Multiasistencia decided to make customer satisfaction a strategic priority. The CEO articulated it thus: "I want to take a qualitative leap in quality. I want to make it a differentiating factor. Today we are the best but we are not rewarded for that because the industry standard is a satisfied/not satisfied binary."

Contracts with insurance companies had traditionally specified target levels of customer satisfaction that were measured at the client level by surveying a sample of policyholders with repairs each month. The specific measure of customer satisfaction and the size of the surveyed sample varied from contract to contract. However, as the CEO noted at the time: "We do not have enough surveys to obtain a precise measure. If we could get a larger sample, and hence a more precise measure of each professional's performance in customer satisfaction, then we could give more weight to the outcome of satisfaction and less to the process metrics relative to what we are doing today." Thus, the firm decided to form a dedicated phone bank with CSRs who

[7] Some of the operating indicators followed were: repair time, use of the PDA to update the state of repair, percentage of customer complaint calls, and percentage of visits resulting in denial of coverage.

would perform the closing call for each repair and, at the same time, survey customer satisfaction.

Multiasistencia wanted a simple customer satisfaction metric that could be incorporated easily into a formulaic bonus plan. They decided to use a simplified version of the Net Promoter Score (NPS) metric.[8] The premise of NPS is that the best way to elicit a sincere and consistent response about the consumption experience is to ask customers whether they would refer the firm to others. The NPS creators believe that a customer makes a personal referral only when they believe the company offers a superior value and understands them. Thus, to assess the customer experience they ask: "On a scale of 0 to 10, how likely is it that you would recommend Company X to a friend or colleague? (0 = never; 10 = very likely)" (Reichheld 2003). Then, they classify customers as *promoters* (score 9–10) who loyally buy from the company and urge their friends to do so, *passives* (score 7–8) who are satisfied but unenthusiastic, and *detractors* (score 0–6) who would avoid any interaction with the company if they could. Multiasistencia decided to use the percentage of detractors among the customers surveyed in a month as the relevant metric for customer satisfaction.

To qualify for the bonus plan in any given month, a professional had to have zero customer complaints.[9] The bonus plan included three performance metrics: the number of detractors, the percentage of repairs fully scheduled with the PDA, and the percentage of repairs that ended in the standard time allotted for that type of job. Repair professionals received 0.70 euros per repair for each of the metrics in which their performance met or exceeded the respective targets. The

[8] NPS is a trademark of Satmetrix Systems Inc., Bain & Co., and Frederick Reichheld.
[9] To count against a professional, the customer complaint had to be based on bad service quality; complaints about denial of coverage were excluded.

targets were set by the management team and considered past performance in different repair specialties.

These targets were:

- 100% of repairs fully scheduled with the PDA
- 80% of repairs ended on time
- 0, 1, or 2 maximum detractors for professionals with less than 30, between 30 and 60, or more than 60 repairs in that month, respectively.

The customer satisfaction phone bank started to formally track customer satisfaction at the professional level in January 2013. Multiasistencia planned to use the data for the period January–March 2013 to help management understand the behavior of the metric. During this period, the information was shared across the management group but not with the repair professionals. In April, regional managers presented the detractors metric and the new bonus system to the repair professionals. The professionals learned about their performance for April via an email at the end of the month.

IV. Experiment Design

Our experiment immediately followed the events described above. Each of the professionals working for Multiasistencia was randomly allocated to one of four treatment groups that received different forms of feedback for a three-month period (May–July 2013). We manipulated two dimensions of that feedback: its frequency and level of detail. Professionals received feedback either on a monthly (M) or weekly (W) basis. Moreover, the feedback was either aggregate (A) or detailed (D). The combination of the two dimensions led to four treatments: *MA*, *MD*, *WA*, and *WD*. At the aggregate level, workers received only information about the total number of detractors during the reporting period. In the detailed treatments, workers received a list of the services with a detractor score (0–6) for the services they finished within the reporting period. The level of detail of the operating performance metrics did not change across treatments and professionals were informed of their percentage use of the PDA and percentage of services closed on time during the reporting period (week or month). Moreover, all professionals also received aggregate measures at the end of the month (as those were the basis for the bonuses they received). The four treatment groups are described in Figure 1.

The experiment began in late April when the company informed the professionals via e-mail of the new feedback protocol. This e-mail was tailored to the specific random assignment of each professional. Professionals were unaware that other types of feedback were provided to other individuals.[10] During the experiment, all professionals were informed about their performance according to their treatment condition.

After the initial information report at the end of April (the monthly aggregate report), which was common to all groups, those in the weekly information cycle received their first performance communication on May 6. Those in the monthly information cycle received their first performance communication on June 3. Subsequently, performance communications were issued on Mondays (for professionals in the weekly cycle every Monday, and for professionals in the monthly cycle on the first Monday after the end of the month).

[10] Professionals worked independently. Even jobs that required the input of multiple professionals (for instance, a broken pipe may have involved the work of both a plumber and a painter) did not require them to work simultaneously, and professionals rarely worked in the same location. Furthermore, the professionals were not unionized. For these reasons, information sharing among professionals was not common. We confirmed the lack of interaction among professionals in the pre-experiment survey. Although some sharing may still have occurred through informal networks, the short time frame of the experiment makes this possibility unlikely.

Prior to the experiment, the company managers strongly preferred to provide more timely information. Because of this, we were not allowed to have a balanced sample in all four treatments. Instead, 25% (75%) of the professionals received monthly (weekly) feedback, and 50% received aggregate or detailed performance information. Thus, we were left with about 100 professionals in treatments *MA* and *MD*, and 300 professionals in treatments *WA* and *WD* (see Table 1).

Because the company started monitoring customer satisfaction in January of 2013, we had four months of data available prior to the experiment.[11] Furthermore, we also administered a questionnaire one year prior to the beginning of the experiment to capture various characteristics of the professionals and to evaluate the risk of spillover across treatments inherent to an individual-level randomization.[12] In addition, we observed four months of post-experiment performance. Figure 2 shows a detailed timeline of the field experiment.

V. Hypotheses

To understand the effects of feedback on performance, we develop a simple model that highlights the value of information for the different treatments.

A Simple Model: The Professional's Decision

Suppose that a professional performs a number of repair services for different customers as assigned by the firm that receives the customer requests. The firm offers the professional implicit and explicit incentives linked to the professional's score on a specific performance metric obtained on those services over a period of time. In our setting implicit incentives come

[11] Multiasistencia started computing this metric early in order to guarantee its consistency before introducing it into the incentive system.
[12] Because the pre-experiment survey was run so far in advance, we believe that it did not contaminate our results.

from the scheduling of future work, and explicit incentives from the bonus linked to monthly performance in customer satisfaction.[13] The monthly performance is a composite metric of the scores obtained in each of the individual services performed by the professional. Further suppose that the individual service scores are independent, i.e., the score a customer gives to the professional after any given service has no impact on the score given by another customer in any other service performed by the professional. Because each service score is independent, a professional maximizes the composite score for the period by maximizing the performance in every individual service i conditional on the information available at every point in time. Thus, if we denote $v(a_i, \theta)$ as the value obtained by the customer when the professional performs service i, the decision problem of the professional is to choose the action $a_i \in A$ that maximizes the expected customer value given the unknown preferences of the customer θ. This formulation ignores agency problems between the firm and the professional for expositional simplicity.

We can allow v to be any arbitrary function, but we assume that $\partial^2 v / \partial a_i \partial \theta \neq 0$ for some (a_i, θ) so that the optimal action a_i depends on θ. We can think of θ as a parameter that captures customer preferences and determines the desirability of different actions, where both θ and the action space A are potentially multidimensional. For instance, θ can measure the typical customer preference for a timely repair, but also the preference of young educated customers for detailed explanations of the repair process, or the preference of older customers for casual conversation. The professional should therefore make a choice as to what dimensions of the action to favor—for instance, either finishing the job as quickly as possible, or, at the risk of extending the length of the service, engaging in technical or social conversation with the customer. The more generally applicable dimensions of θ such as the preference for timeliness of

[13] To simplify the exposition we assume that the conflicts between objectives that occur in a multitasking setting do not apply here.

service may be identified through statistical analysis by a centralized unit. However, the professional may improve on the broad directives from the centralized unit by exploiting the soft information captured through the interaction with the customers.

We assume that θ is unknown and unobservable to the professional, but he learns about the likely preferences of the customer through information provided by performance feedback. Specifically, by observing customer satisfaction scores obtained in past service events and linking them to action choices and to observed customer behaviors and characteristics, the professional may infer θ with more or less precision.

Every time the professional performs a service a score s_i is generated to evaluate the professional's actions. We assume that past performance provides a set of signals $s_{t,i}$, corresponding to each of the services $i = 1, \dots, k$ performed in period t, that are informative about θ. By observing the scores received in past services and reflecting on the specifics of the intervention, the professional may infer on the adequacy of certain action choices to serve customers with some observed characteristics. The signals are informative about θ because they reflect customer preferences, but in a noisy way (for instance, because the survey evaluation may be affected by the mood of the customer the day he/she is questioned, or because the customer has idiosyncratic preferences that depart from average needs of the population). We further assume that all past signals are equally informative about θ.[14]

We take the period $t = (m, w)$ to correspond to a month m and a week w, and we assume four weeks in one month, i.e. $w \in \{1,2,3,4\}$. Because not all professionals observe the

[14] This assumption implies that customer preferences are constant over time, so that θ affects the services in all time periods equally. As a result, the most recent services are as informative as older services. Because the time period we consider in our study is relatively short, customer preferences are unlikely to change significantly.

same performance history, we denote by I_t^F the information available to the professional at time *t*. Then, his objective when performing a service *i* at time *t* is

$$\max_{a_i \in A} E(v(a_i, \theta) | I_t^F),$$

where the information available to the professional for making the decision depends on the feedback treatment $F \in \{MA, MD, WA, WD\}$.

The Case of a Bayesian Professional

A Bayesian professional uses all available information in a rational way, recalling all past feedback reports (Savage 1954). Therefore, such a professional under the MA treatment observes the average scores for each and all of the preceding months, i.e. $I_{(m,w)}^{MA} = \{\bar{s}_n\}_{n<m}$, where $\bar{s}_n = \frac{1}{4k}\sum_{x,i} s_{(n,x),i}$ is the average performance for month n.[15] A professional under MD treatment observes the individual score for all the services performed in previous months but does not observe performance for the current month, i.e. $I_{(m,w)}^{MD} = \{s_{(n,x),i}\}_{n<m,x,i}$. Under the WA treatment, he observes the average signal for each of the preceding weeks (including the previous weeks of the current month), i.e. $I_{(m,w)}^{WA} = \{\bar{s}_{(n,x)}\}_{n<m,x} \cup \{\bar{s}_{(m,x)}\}_{x<w}$, where $\bar{s}_{(n,x)} = \frac{1}{k}\sum_i s_{(n,x),i}$ is the average performance for week x in month n. Finally, a professional under WD treatment observes the detailed signal of all the services performed up to the current week, i.e. he has the same information as the professional in the *MD* treatment, plus he observes the previous weeks of the current month, so that $I_{(m,w)}^{WD} = I_{(m,w)}^{MD} \cup \{s_{(m,x),i}\}_{x<w,i}$.

[15] In our setting, because the professional knows the number of services, observing the average signal is equivalent to observing the number of detractors.

Because all signals $s_{t,i}$ are equally informative, the more signals the professional observes, the more precise the inference about θ and, as a result, the higher the expected value obtained from the choice of action. This allows us to rank the expected performance of a professional under the different feedback regimes. Because the *WD* treatment provides at least as many signals as the *MD*,[16] it should lead to better performance. Similarly, the average monthly signal $\bar{s}_n$ is less informative than observing each of the weekly averages $\bar{s}_{(n,w)}$ for $w = 1, \dots, 4$.[17] As a result, the *WA* treatment should lead to better performance than the *MA* treatment. Furthermore, because both the monthly and weekly averages contain less information than the detailed signals of the same time period, we should observe better performance from *WD* than *WA* and from *MD* than *MA*. Finally, notice that in general it is not possible to rank the information content of $I_{(m,w)}^{WA}$ and $I_{(m,w)}^{MD}$ because, although treatment MD has more detailed information, the WA treatment starts receiving feedback about earlier weeks in the current month m while treatment MD does not. However, during the first week of the month, the MD treatment has more information than the WA treatment (they both have information about the same time periods, but the information is more detailed for the first treatment).[18]

[16] Although the WD treatment receives more updates than *MD* (and hence has a more informative signal) during the later weeks of the month, the information content of both treatments is the same during the first week of the month, as both observe detailed performance on all past services.

[17] Note that the information set of the *WA* treatment includes the set of the *MA* treatment. Knowledge of the number of services performed every week allows the professional to calculate the average monthly score from the average weekly scores.

[18] If we compare a professional in each of the four treatments with the information derived from the same realizations of the underlying signals, the result follows immediately. To see this, notice that both $\bar{s}_n$ and $\bar{s}_{(n,w)}$ are a function of the underlying signals $s_{t,i}$ for that period, and $\bar{s}_n$ is a function of $\bar{s}_{(n,w)}$ for $w = 1, \dots, 4$. As a result, any action profile $a^*(\bar{s}_n)$ that is optimal under treatment *MA* can be replicated by any of the other treatments (by simply disregarding the additional information they have), with similar arguments applying to the other comparisons described above.

However, the result can be generalized to the case in which the information of the four treatments is drawn from the same underlying distribution of the signals, but not necessarily from the same realizations (for instance, because different professionals observe the realization of the score for different jobs, all of which are randomly drawn from the same population of potential customers). In this case, we can rank the informativeness of the signals using Blackwell's order, denoted $\succcurlyeq$. Then, Blackwell's theorem implies that $I_{(m,w)}^{MA} \preccurlyeq I_{(m,w)}^{MD}, I_{(m,w)}^{WA} \preccurlyeq I_{(m,w)}^{WD}$ for all m, w.

If we denote the expected performance of a professional under treatment F for a service *i* at time t by $Ev_t^F = \max_{a_i \in A} E(v(a_i, \theta) | I_t^F)$, then we get the following result:[19]

PROPOSITION 1. *The expected performance of a Bayesian professional with perfect recall and unbounded rationality satisfies:*

1. $Ev_t^{MA} \leq Ev_t^{MD}, Ev_t^{WA} \leq Ev_t^{WD}$ *for all* t.
2. $Ev_t^{WA} \leq Ev_t^{MD}$ *for* $t = (m, 1)$.
3. $Ev_t^{MD} = Ev_t^{WD}$ *for* $t = (m, 1)$.

The result shows that more information is always better for a professional who conforms to the model of the Bayesian decision maker, and therefore feedback is most effective when it is both detailed and frequent.

The Case of a Local Thinker Professional

Suppose now that the professional is not perfectly Bayesian. In particular, we will assume that the professional overweights the last report when making inferences about the right course of action. In Gennaioli and Shleifer's (2010) terminology, the last feedback report is salient and the professional is a local thinker. Although salience may be the result of an irrational bias in the cognitive processing of the information, it is also possible that overweighting recent information is the rational response to the cognitive limitations of professionals, such as the inability to perfectly recall past reports, or the presence of sizeable costs of retrieving past feedback

Furthermore, during the first week we have $I_{(m,1)}^{WA} \preccurlyeq I_{(m,1)}^{MD}$ and $I_{(m,1)}^{MD} \sim I_{(m,1)}^{WD}$. Proposition 1 then follows from these relations (Blackwell 1951).

[19] Note that the extension of the results to the set of services performed during the bonus period is trivial because the period (monthly) score is simply an equally weighted linear combination of individual service scores.

reports.[20] We are neutral on these alternative interpretations, as our research design does not allow us to distinguish among them. Nevertheless, it is still the case that the way information is presented affects its salience, and hence its use, and that feedback systems could be designed to minimize this effect.

To simplify matters, we will assume that the professional uses only the information contained in the last feedback report, disregarding all previous information, but the results hold if we assume other patterns of decay in the use of information. Hence, a local thinker professional observes the monthly average for the previous month under the MA treatment, i.e. $I_{(m,w)}^{L,MA} = \{\bar{s}_{m-1}\}$; the individual signals for the previous month under the MD treatment, i.e. $I_{(m,w)}^{L,MD} = \{s_{(m-1,x),i}\}_{x,i}$; the average for the previous week under the WA treatment, i.e. $I_{(m,w)}^{L,WA} = \{\bar{s}_{(m-1,4)}\}$ if $w = 1$ or $I_{(m,w)}^{L,WA} = \bar{s}_{(m,w-1)}$ if $w > 1$; and the individual signals for the previous week under the WD treatment, i.e. $I_{(m,w)}^{L,WD} = \{s_{(m-1,4),i}\}_i$ if $w = 1$ and $I_{(m,w)}^{L,WD} = \{s_{(m,w-1),i}\}_i$ if $w > 1$.

Because a local thinker disregards past feedback, the informativeness of the feedback treatments reverses. At a given level of detail, the professional receiving more frequent feedback disregards more information. As a result, the signal(s) he uses is less informative, as it contains information for one week instead of one month.[21] As before, however, it is not possible to compare the signals that result from changing both the level of detail and the frequency. In this

[20] Because feedback reports do not contain performance data for earlier periods, accessing that information would require the time commitment to retrieve past emails containing the feedback reports, and an ex-ante effort to store and perhaps organize those emails for easy future access. Professionals may have deemed those costs excessive relative to the benefits of having such information easily accessible.
[21] Notice that this result rests on the assumption of the informative equivalence of all past signals, i.e., that no signal, including a more recent signal, is more informative than any other signal.

case, the WD treatment contains more detailed information, but over fewer services than the MA treatment. As a result, it is not possible to rank the two treatments.[22]

We obtain the following result for the expected performance of such a professional:

PROPOSITION 2. *The expected performance of a local thinker professional satisfies:*

1. $Ev_t^{WA} \leq Ev_t^{MA}, Ev_t^{WD} \leq Ev_t^{MD}$ *for all* t.

The result highlights the fact that more information is not always better when the professional is a local thinker. The way the information is presented affects the professional's ability to process it. In this case, we are likely to see the best results from feedback information that is detailed but infrequent.

Along with the customer satisfaction metric, Multiasistencia provides feedback on two other process metrics: services finished on time and interventions scheduled through the PDA application. The professional cannot be certain about the rating the customer will provide. In that sense, feedback on customer satisfaction provides new performance information. In contrast, performance in both of the process metrics is evident immediately, as the professional knows whether she schedules a job through the PDA or whether she finishes a job on time before she receives official feedback from the firm. Therefore, there is no new information in the feedback communicated to the professional for these metrics. If the signal is uninformative, $E(v(a_t,\theta)|I_t^T) = E\big(v(a_t,\theta)\big)$ and hence the professional can achieve the same expected performance under all treatments.

PROPOSITION 3. *If the feedback is uninformative, then the expected performance is the same for all treatments regardless of whether the professional is Bayesian or a local thinker.*

[22] Hence, using Blackwell's order, we have $I_{(m,w)}^{WA} \preccurlyeq I_{(m,w)}^{MA}, I_{(m,w)}^{WD} \preccurlyeq I_{(m,w)}^{MD}$ (see footnote 18).

In the next section, we discuss how the data can shed light on the importance of these effects for the effective release of feedback information.

VI. Experiment Results

In this section we compare the performance of professionals in the four treatments to identify the value of frequent and detailed feedback.

Summary statistics, presented in Table 1 Panel A, show that professionals' performance in the three metrics of interest is similar across all treatments in the first four months of 2013, suggesting a successful randomization. This is corroborated in Table 1 Panel B in a regression framework: only the coefficient on the *MD* treatment in the PDA model is statistically significant at the 10% level, but the F-test fails to reject that the coefficients on the three treatment dummies are jointly insignificant.

The first result of the paper can be seen in Table 1 Panel A. It shows the average share of detractors for each of the four treatments. Professionals in all four treatments improve their performance (fewer detractors) between the pre-experiment and experiment periods, an effect that may be due to the introduction of the incentives, the introduction of the feedback, or a combination of both.

The three months of the experiment show the control group (*MA*) performing just as well as the weekly treatments (*WD* and *WA*), while professionals in the treatment *MD* show the most improvement in performance, achieving the lowest share of detractors of the four groups (8.37%).

A similar picture emerges when we look at the fraction of professionals with zero detractors in a month. This fraction increases for all groups during the experiment months, but it does so more markedly for treatment *MD* than for the others.

We also observe an improvement in the operational metrics included in the bonus program during the experiment period, but the improvement is very similar across all treatments.

We develop these insights below, with additional statistical analyses.

i. The effects of the amount and frequency of feedback

To formalize our inference about the treatment effects and assess the robustness of our results, we estimate various regression models. Because professionals are randomly assigned to one of the four treatments, we can estimate average treatment effects by comparing the average performance of the professionals assigned to each treatment during the three-month experiment period with the following regression:

$$y_{it} = \beta_0 + F_i\beta_1 + X_{it}\delta + \varepsilon_{it} \tag{1}$$

where y_{it} is the performance of professional i in period t, F_i is a vector of feedback treatment indicators, and X is a vector of additional covariates. The controls in X include time effects, to control for time trends, and the repair specialty of the professional, to account for heterogeneity in professionals' characteristics. In the regressions, we drop the dummy for the control treatment (*MA*) so that the constant captures the average performance for this group and the coefficients on the other three treatment dummies measure the difference in performance relative to the control.

We begin by looking at the performance of professionals delivering customer satisfaction, as measured by the share of detractors. The estimates without controls presented in column (1) of

Table 2 show that the professionals in treatment *MD* perform better than those in the control group (*MA*). They manage to lower their share of detractors to 2 percentage points below professionals in the *MA* treatment. This difference represents a sizeable 20% improvement relative to the 10% share of detractors in the control group. However, the professionals in the two weekly treatments (WA and WD) show no difference in performance with respect to the control group. The same results follow when we control in column (2) for month effects and for the specialty of the professional.

Because we also observe the professionals for the four months prior to the experiment, we compare the improvement in performance between the three months of the experiment and the previous four months for the four treatments using a difference-in-differences estimation. In this way we control for any heterogeneity across treatment groups that could have arisen spuriously during the random assignment process. We do so by including the vector of treatment indicators F_i, a dummy *D* indicating the treatment period, and their interaction in the following linear model:

$$y_{it} = \beta_0 + F_i\beta_1 + D_t\beta_2 + D_tF_i\beta_3 + X_{it}\delta + \varepsilon_{it} \qquad (2)$$

where the vector of covariates *X* now includes not only time and specialty effects, but also individual fixed effects to control for any unobserved heterogeneity. As before, we also drop the dummy for the control treatment, so that the interaction terms capture the performance of the other three treatments relative to the control group.[23]

The estimates in model (3) of Table 2 show the basic difference-in-differences estimation, without any controls. We can see that the overall share of detractors is lower in the three months

[23] These regression models are estimated with a total of 4,722 observations, which correspond to the sum of observations prior to and during the experiment (2,589 and 2,133, respectively).

of the experiment than in earlier months, showing that performance improves after the introduction of the bonus and feedback system. Moreover, the professionals in treatment *MD* improve performance by more than the control group (*MA*). They manage to lower their share of detractors by 3.4 percentage points more than the 2.5 percentage points drop observed in the control group (representing 46% and 19% improvement respectively relative to the baseline 13% of detractors). The professionals in the two weekly treatments (*WA* and *WD*) also have fewer detractors, but their decrease is not statistically different from that in the control group.

Adding the pre-treatment period performance also allows us to perform the stricter control for unobserved heterogeneity using individual (professional) fixed effects. Column (4) reports the results with monthly and individual effects. We obtain the same results, reassuring us that they are not due to an imperfect random assignment of professionals to treatments. Professionals in treatment *MD* improve their performance relative to the control group, but the weekly treatments are indistinguishable from the control.

Because we have a large number of observations with zero detractors (see Table 1), we also estimate a Tobit model in columns (5) to (8) to make sure that the zero bound is not driving our results even if we can no longer interpret the coefficients as a percentage of detractors.[24] As expected, the coefficients from the Tobit model are larger (in absolute terms) than those of the linear probability model (OLS). Nonetheless, we find the same results. Professionals in the *MD* treatment perform significantly better than those in the control group, while the performance of professionals in the monthly treatments (*WA* and *WD*) and the control are indistinguishable.[25]

[24] In this case, however, we do not have individual effects because the maximum likelihood estimator of the Tobit model is inconsistent under fixed effects. We use specialty effects instead.

[25] The F-tests reported in Table 2 further support our conclusions. They generally reject the hypothesis that all three treatment coefficients are jointly insignificantly different from zero.

Next, we turn to two alternative measures of customer satisfaction: the proportion of promoters and the average survey score. Because the proportion of observations with extreme values (0 or 100% of promoters, and 0 or 10 score) is very small, we only report the OLS results.[26] Columns (9) and (10) provide the results for promoters, controlling for month and professional fixed effects. Although the share of promoters increases over time, we find no differential effect for any of the treatments. Because the number of promoters does not affect professionals' compensation—but the number of detractors does—the professionals probably concentrate their efforts on using feedback to improve their performance in the most difficult services (the ones that were likely to yield a low value in the survey).

If we compare the average score in the customer satisfaction survey (columns (11) and (12)), we again find that treatment *MD* is the only one that improves upon the control group. However, because the improvement in performance only happens for a fraction of the services provided by this group, the economic effect is smaller than in the results described above. The average score is 8.0 in the first four months of 2013. This score increases by almost half a point (or about 6%) during the experiment period for the control group, and increases by an additional 0.2 points (or 2.5%) for professionals in treatment *MD*.

The results presented in this section suggest that providing more detailed feedback is useful for improving performance. However, that is only the case when feedback is provided sparsely. Detailed feedback loses its usefulness when provided very frequently. In fact, the F-test shows that the effect of *MD* is significantly different from that of *WD*, suggesting that performance deteriorates when detailed information is provided more frequently. Similarly, providing more

[26] The results from the Tobit model are essentially the same, and are available from the authors upon request.

frequent feedback, even when it is less detailed, does not seem to help professionals improve their performance.[27]

Taken together, the results suggest that professionals fail to fully process all available detailed information when it is provided frequently. The recipient of frequent feedback may fixate on the most recent and salient information, leading him or her to underweight or ignore evidence that is more distant in time and thus limiting the amount of information actually used in decision-making. This leads professionals to make the wrong inferences, reducing their learning and hampering performance improvement. By providing detailed but less frequent feedback, Multiasistencia communicates richer information in a single report, allowing professionals to identify true trends and ignore noise in the metric.

ii. Feedback on customer satisfaction vs. operational performance

We now turn to the effects of feedback on the two measures of operational performance that are also part of the incentive scheme: scheduling services via the PDA and completing them on time. These two measures are of a very different nature than customer satisfaction: they capture the input of the professionals, while customer satisfaction is a measure of their output. Professionals can perfectly observe the performance of the former directly but they do not observe the latter until they receive feedback from the firm. Because the feedback does not provide any additional information on the operational performance measures, we should not expect to find differential effects for the different feedback treatments. The only possible

[27] In untabulated analyses we test whether there are heterogeneous responses to the treatment effects we identify. To this end, we consider several measures of professionals' ability and experience (past performance, tenure, and education). The results show no heterogeneous treatment effects. Therefore, the effects of the detailed and infrequent feedback seem to be widespread and fairly homogeneous.

exception would be if the feedback acts as a reminder that those dimensions of performance are important for the firm's management.

Table 3 estimates analogous models to those in Table 2 using Tobit, where the dependent variable is either the share of services closed on time or the share of services that were properly scheduled using the PDA.[28] As expected, the results show no differences among the treatments on these two dimensions. Not only are the coefficients statistically insignificant, but the magnitudes of the effects are also economically negligible.

iii. Salient feedback vs. dynamic incentives

The evidence presented so far is consistent with the hypothesis that more information about output-based performance measures is useful when it is provided within a timeframe that allows professionals to make meaningful inferences from it. The same information becomes less useful when it is provided too frequently, as past feedback is disregarded. However, the evidence is also consistent with the presence of dynamic incentives, as professionals under frequent feedback conditions may perform worse because they learn earlier that they will not meet the monthly goal. Thus, they may lower their effort in the later part of the month.[29]

In this section we provide further evidence that is consistent with professionals being local thinkers reacting to salient (more recent) information by disaggregating performance at the weekly level. Our results are consistent with professionals acting on inferences based on partial information when feedback is frequent (weekly). This evidence is unlikely to be explained by dynamic incentive considerations.

[28] We do not show the estimates from the OLS models, although the results remain the same, and are available upon request.
[29] Non-linear incentive schemes are known to create dynamic incentives, with varied responses over time based on past performance (see, for instance, Casas-Arce and Martínez-Jerez 2009).

Table 4 provides Tobit estimates of treatment effects using weekly data. Column (1) simply estimates average treatment effects controlling for specialty and time (weekly) effects. As with column (8) of Table 2, we find that professionals in monthly treatment *MD* improve their performance relative to the control group, while the weekly treatments *WA* and *WD* show no improvement (the coefficient on *MD* has a p-value of 0.15 and the F-test strongly rejects the possibility that the three treatment indicators are jointly insignificant).[30]

Next, we separate the treatment effects for the first week (before the weekly treatments receive any feedback about the current month) and for the rest of the month by estimating the following model:

$$y_{it} = \beta_0 + F_i\beta_1 + D_t^e\beta_2 + D_t^l\beta_3 + D_t^eF_i\beta_e + D_t^lF_i\beta_l + X_{it}\delta + \varepsilon_{it} \qquad (3)$$

where t now denotes weeks rather than months, D^e takes a value of 1 for the early part of the treatment months (the first week of each month) and 0 otherwise, and D^l takes a value of 1 for the later part of the treatment months. Because we omit the dummy for treatment *MA*, the coefficients on the other three treatments show their performance during the period relative to the control group.

The results in column (2) show that treatments *WA* and *WD* do just as well as the control group in the later part of the month, while they seem to perform worse in the first week (the coefficient is statistically significant and of similar size for both *WA* and *WD*). Furthermore, the weekly treatments do worse than treatment *MD* both in the first week and in the later part of the month. This result is inconsistent with a Bayesian thinker (proposition 1) but it is not completely

[30] In fact, the size of the *MD* treatment coefficient is the same as that in column (8) of Table 2. However, the standard error is larger due to the higher variance in the weekly data.

explained by dynamic incentives either. If the lower performance in the weekly feedback treatments were due exclusively to professionals rationally abandoning the pursuit of their bonus targets after receiving bad news, we would expect to observe statistically indistinguishable performance across all treatments in the first week of the month and deteriorated performance in the later weeks of the month for the weekly treatments.

To further explore this issue, we first split the effect of the second part of the month for those professionals with at least one detractor in the first week of the month from those with none. Model (3) in Table 4 estimates the following regression:

$$y_{it} = \beta_0 + F_i\beta_1 + D_t^e\beta_2 + D_t^l NoDetr_{it}\beta_3 + D_t^l Detr_{it}\beta_4 + D_t^e F_i\beta_e + D_t^l NoDetr_{it}F_i\beta_{l,NoDetr} + D_t^l Detr_{it}F_i\beta_{l,Detr} + X_{it}\delta + \varepsilon_{it} \quad (4)$$

where *NoDetr* is an indicator function that takes a value of 1 if the professional does not receive any detractors in the first week of the month (and hence still qualifies for the bonus), while *Detr* indicates that there is at least one detractor in the first week.

The results in column (3) show that professionals in treatments *WA* and *WD* who receive at least one detractor in the first week of the month (and learn about it through their weekly feedback) perform significantly worse in the second part of the month than professionals in the control group and in the *MD* treatment who also receive a detractor the first week (but are unaware of it until the end of the month). Thus, professionals in the weekly treatments under-perform after receiving bad news. This evidence is consistent with the presence of dynamic incentives (Casas-Arce and Martínez-Jerez 2009). Because the bonus is paid monthly, the professionals in weekly treatments learn their interim performance and can adjust their effort accordingly. These professionals may still process the information efficiently, but may fail to improve their performance because they know they have been disqualified for a bonus well before the end of the month. In fact, if this is the case they may rationally abandon their pursuit of the bonus and lower their effort in the final weeks.

However, professionals in treatment *MD* significantly outperform the control group and the two weekly feedback treatments in the second part of the month if they do not receive a detractor in the first week. The F-test rejects at the 1% level the equality of the coefficients on *MD* and *WD* treatments (and at the 5% for the *MD* and *WA* treatments). In this case there is no change in the *WD* and *WA* professionals' expectations of achieving the bonus that justifies their exerting a lower effort than professionals in the *MD* treatment. Thus, this pattern of performance—expected if professionals are local thinkers—cannot be explained by the rational abandonment of the bonus.

Notice also that dynamic considerations should only affect performance at the end of the month. Because the performance measure is reset each month for bonus calculation purposes, we should observe no difference in performance in the first week of the month based on performance the week before if the results for weekly treatments are driven by dynamic incentives. However, we could still see a negative effect if the professionals are reacting to frequent feedback information for other reasons.

To test whether dynamic effects explain the professionals' behavior, we next separate the treatment effect for the first week of the month for those who receive at least one detractor in the preceding week (the last week of the previous month) and those who do not by estimating the following model:

$$y_{it} = \beta_0 + F_i\beta_1 + D_t^e NoDetr_{it}\beta_2 + D_t^e Detr_{it}\beta_3 + D_t^l\beta_4 + D_t^e NoDetr_{it}F_i\beta_{e,NoDetr} + D_t^e Detr_{it}F_i\beta_{e,Detr} + D_t^l F_i\beta_l + X_{it}\delta + \varepsilon_{it} \qquad (5)$$

where *NoDetr* now indicates that there is no detractor in the last week of the previous month, while *Detr* indicates that there is a detractor.

The estimates are in column (4) of Table 4. The results for the first week are similar to those for the later part of the month (if anything, they are even stronger), suggesting that dynamic incentives are not the main cause. Most striking is the fact that the *WD* treatment performs worse than the *MD* in the first week of the month after receiving feedback of a negative outcome, despite the fact that both groups have the same amount of information. This evidence is inconsistent with the presence of dynamic incentives and strongly suggests that the last feedback report is most salient for all professionals.

However, from table 4 it is difficult to identify whether salience leads professionals to make incorrect inferences by overweighting the importance of any detractors in the previous week or to suffer a temporary loss of confidence after receiving negative feedback (McCarty 1986). To tease apart the two alternative explanations for the salience effect of the last feedback report, we look at the performance variance after a negative outcome. If professionals in the weekly treatment suffer a temporary loss of confidence, this will translate into a loss of motivation, reduction of effort, and deterioration of performance. Thus, both the mean and the variance of the professionals' performance will decrease.[31] In contrast, if those professionals infer a course of action from incomplete information, sometimes they will err and their performance will

[31] Notice that the same would happen if the professionals lower their effort after rationally abandoning their pursuit of the bonus.

deteriorate but other times they will be correct and their performance will improve. Thus, their mean performance will deteriorate while their variance will increase.

Table 5 compares the performance variance of professionals in the weekly and monthly treatments after receiving positive and negative feedback. The performance variance of professionals in the weekly treatments is generally higher than that of professionals in the monthly treatments and increases after a negative outcome, consistent with professionals being local thinkers. That is, we find evidence consistent with our results being driven mainly by professionals making inferences not using all available information and not by a temporary loss of confidence induced by a negative feedback report.

In summary, we cannot completely rule out the possibility that professionals in weekly treatments also rationally lower their effort when the negative feedback implies that they will not reach their bonus target, i.e. that they respond to dynamic incentives. Nevertheless, the evidence presented in this section suggests that the main cause of the effects we observe is salience—professionals are local thinkers and overweight the information contained in the last feedback report.

iv. Post-experiment performance

Finally, we look at the post-experiment performance of the professionals in the different treatments. The initial assumption of Multiasistencia's management team was that the professionals with access to more and more frequent information would perform better. Their prior was stronger with respect to frequency, to the point that they required us to place many more professionals in the weekly treatment groups than in the monthly treatments (300 each versus 100 each). However, as the experiment unfolded, we provided them with preliminary evidence of the superiority of the *MD* treatment. In view of these results, Multiasistencia decided to provide all professionals with monthly and detailed detractor information after the experiment ended.

Once all the professionals receive feedback with the same level of detail and the same frequency, we expect to observe similar patterns of performance throughout the firm. To test this, we collect information on the four months following the experiment phase (August to November) to provide some additional robustness tests.

Table 6 presents the results. In column (1) we estimate the same Tobit model as in column (8) of Table 2 (difference-in-differences with specialty and time effects), but augmented with a dummy for the post-experiment period interacted with the treatment groups.[32] (We do not report the treatment effects during the experiment months, as they are analogous to those in Table 2.) The results show that the improved performance with respect to the pre-experiment period persists in the post-experiment period. However, the coefficients on the treatment dummies for the post-experiment period are individually and jointly insignificantly different from zero, indicating that now there are no differences among the treatment groups. That is, the performance advantage of the *MD* treatment disappears as soon as the other professionals start receiving the same feedback treatment.

Column (2) then extends the Tobit model in column (4) of Table 4 to look at the differential reaction contingent on past performance (as in column (1), we do not report the coefficients for the experiment period, which are analogous to those in Table 4). The results show that as soon as professionals in the *WA* and *WD* treatments stop receiving weekly information, their

[32] Notice that the experiment period dummy takes a value of 1 only from May to July. Therefore, the coefficient of the post-experiment dummy measures the change in performance relative to the pre-experiment period (January to April), rather than relative to the experiment period.

performance is indistinguishable from that of the professionals in the monthly treatments regardless of the week of the month and of whether they received a detractor in the prior week.

These results provide further evidence suggesting that the experiment results were not the product of chance. As soon as professionals start receiving feedback information with the same frequency and detail, their performance differences vanish.[33] In this regard, the fact that the *MD* treatment loses its advantage relative to the other treatments suggests that the effects of information are short-lived. This result is consistent with the assumption that professionals disregard past information in light of the latest feedback report, which is most salient to them.

VII. Conclusion

This paper presents evidence on how the characteristics of information drive improvements in performance by decision-making employees. Using a field experiment that manipulates the frequency and detail of the non-financial performance feedback received by professionals in a property repair company, we find that detailed information leads to a significant improvement in performance. However, contrary to what we would expect if professionals used all the information available, detailed information is only useful when provided sufficiently sparsely. When feedback is too frequent, professionals perform significantly worse than a group with detailed and less frequent information. This evidence is consistent with decision makers (the repair professionals) not being able to properly process detailed information when it is provided too often. Professionals seem to fixate on the information that is more salient (contained in the last feedback report), disregarding past reports.

[33] Incidentally, the post-experiment results also show that the professionals in the weekly treatments (the *WD* treatment in particular) did not withhold performance in the short term during the experiment (for instance, by engaging in more experimentation to increase learning) in order to improve their performance in the long term.

Because our evidence derives from a field experiment that is solidly grounded in theory, we expect the inferences drawn from our results to be widely applicable to other settings. They should be especially relevant in environments where there is significant scope for learning and where employees have discretion about how to direct their efforts.

One design feature of the feedback system in our study that should be generalized with special caution is the relevant frequency range. In our setting, monthly is the natural base for feedback frequency because of the nature of the tasks (finished in a few days and repeated several times a day) and the compensation cycle of the industry. Weekly measurement is thus a logical increase in frequency. In settings with longer or less frequently repeated tasks, the relevant feedback frequency options may differ. In settings where timely information is a key feature (e.g. brokerage firms), our results for feedback frequency may not apply at all. Finally, our results may not be generalizable in settings where feedback is not linked to bonuses. Prior research finds that feedback effects are more consistent when associated with incentives, but this is not always the case (Kluger and DeNisi 1996; Balcazar et al. 1985; Alvero et al. 2001).

However, even in settings with different features from ours, the fact that certain types of information may be salient for decision makers needs to be kept in mind when designing organizations for performance. Advances in technology have facilitated the capture and prompt delivery of performance information within the corporation. In contrast with the common assumption that more and more frequent information always yields better results, our findings suggest that managers should weigh the benefits of detailed, immediately available information against the ability of the recipients to properly process that information. Managers should also pay careful attention to the design of their accounting systems in order to facilitate the efficient processing of information, taking into account the fact that individuals may not make inferences using all available information if some information is delivered in a more salient form. In short, the way information is presented affects how it is consumed and acted upon.

The implications of salience may extend well beyond the managerial accounting domain to the reporting of financial information in capital markets. Although our evidence does not directly speak to the effects of financial reporting, a similar rationale is often cited to justify reductions in reporting frequency. In the United States, McDonald's stopped reporting monthly store sales in June 2015 because, according to its CEO, "the monthly reporting just lends itself to more volatility, and I think investors focus on short-term issues."[34,35] Rational explanations for the effects of reporting frequency have been proposed in the literature (see, for instance, Gigler et al. 2014), but salience is likely to play an important role as well. Exploring these avenues could provide fertile ground for future research.

---

[34] http://www.bloomberg.com/news/articles/2015-05-27/mcdonald-s-to-stop-reporting-monthly-same-store-sales (accessed on June 11, 2015).

[35] Other prominent firms that have made similar reporting choices include restaurant chains such as Yum! Brands Inc., Chipotle Mexican Grill Inc., and Starbucks Corp., as well as retailers such as Macy's and J.C. Penney Co. In the United Kingdom, following the removal of the requirement to publish interim financial statements, companies such as National Grid also stopped reporting quarterly results because they "can frequently provide an unnecessary focus on matters of little relevance to a long-term business" (http://www.ft.com/cms/s/0/3f4a6fd8-a5fc-11e4-abe9-00144feab7de.html#axzz3chCalDPF accessed on June 11, 2015).

References

Alvero, A. M., Bucklin, B. R., and Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985–1998). *Journal of Organizational Behavior Management* 21(1): 3–29.

Ammons, R. B. (1956). Effects of knowledge of performance: A survey and tentative theoretical formulation. *Journal of General Psychology* 54(2): 279–299.

Annett, J. (1969). *Feedback and Human Behaviour*. Harmondsworth, Middlesex, England: Penguin Books.

Balcazar, F., Hopkins, B. L., and Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management* 7(3-4): 65–89.

Bilodeau, E. A. (1969). Supplementary feedback and instructions. In Bilodeau, E. A. (Ed.), *Principles of Skill Acquisition* (pp. 235–253). New York: Academic Press.

Blackwell, D. (1951). Comparison of experiments. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (pp. 390-395). University of California Press.

Bohn, R. (1995). Noise and learning in semiconductor manufacturing. *Management Science* 41(1): 31–42.

Bohn, R. (2000). Stop fighting fires. *Harvard Business Review* 78(4): 82–91.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics* 127(3): 1243–1285.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. *The Journal of Political Economy* 121(5): 803–843.

Bushee, B., and Noe C. (2000). Corporate Disclosure Practices, Institutional Investors, and Stock Return Volatility. *Journal of Accounting Research* 38: 171–202.

Casas-Arce, P., F.A. Martinez-Jerez, and V.G. Narayanan. (2016). The Impact of Forward-Looking Metrics on Employee Decision-Making: The Case of Customer Lifetime Value. Working Paper University of Notre Dame.

Casas-Arce, P., Martinez-Jerez, F. A. (2009). Relative performance compensation, contests, and dynamic incentives. *Management Science* 55(8): 1306–1320.

Chhokar, J. S., and Wallin, J. A. (1984). A field study of the effect of feedback frequency on performance. *Journal of Applied Psychology* 69(3): 524–530.

Feltham, G., and Xie, J. (1994). Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review* 69(3): 429–453.

Floyd, E., and List, J. A. (2016). Using field experiments in accounting and finance. *Journal of Accounting Research* (forthcoming).

Gennaioli, N., and Shleifer, A. (2010). What comes to mind. *The Quarterly Journal of Economics* 125(4): 1399–1433.

Gigler, F., Kanodia, C., Sapra, H., and Venugopalan, R. (2014). How frequent financial reporting can cause managerial short-termism: An analysis of the costs and benefits of increasing reporting frequency. *Journal of Accounting Research* 52(2): 357–387.

Goodman, J. S., Hendricks, M., and Wood, R. E. (2004). Feedback specificity, exploration, and learning. *Journal of Applied Psychology* 89(2): 248–262.

Hannan, R. L., Krishnan, R., and Newman, A. H. (2008). The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. *The Accounting Review* 83(4): 893–913.

Hannan, R. L., McPhee, G. P., Newman, A. H., & Tafkov, I. D. (2013). The effect of relative performance information on performance and effort allocation in a multi-task environment. *The Accounting Review* 88(2): 553–575.

Ilgen, D. R., Fisher, C. D., and Taylor, M. S. (1979). Consequences of individual feedback on behavior in organization. *Journal of Applied Psychology* 64(4): 349–371.

Ilies, R., and Judge, T. A. (2005). Goal regulation across time: The effects of feedback and affect. *Journal of Applied Psychology* 90(3): 453–467.

Ittner, C. D., and Larcker, D. F. (1998). Are non-financial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research*, 36 (Supplement), 1–46.

Kahneman, D., and Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology* 3(3): 430–454.

Kahneman, D., and Tversky, A. (1974). Judgment under uncertainty: Heuristics and biases. *Science* 185(4157): 1124–1131.

Kahneman, D., and Tversky, A. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review* 90(4): 293–315.

Kang, K., Oah, S., and Dickinson, A. M. (2005). The relative effects of different frequencies of feedback on work performance: A simulation. *Journal of Organizational Behavior Management* 23(4), 21–53.

Kiefer, N., and Nyarko, Y. (1995). Savage-Bayesian models of economics. In A. Kirman and M. Salmon (Ed.), *Learning and Rationality in Economics* (pp. 40–62). Oxford, UK: Basil Blackwell Press.

Kim, J. S. (1984). Effect of behavior plus outcome goal setting and feedback on employee satisfaction and performance. *Academy of Management Journal* 27(1): 139–149.

Kluger, A. N., and DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 119(2): 254–284.

Kopelman, R. (1986). Objective feedback. In Locke, E. A. (Ed.), *Generalizing from the Laboratory to Field Settings* (pp. 119–146). Lexington, MA: Lexington Books.

Leskovec, E. W. (1967). A guide for discussing the performance appraisal. *Personnel Journal* 46, 150–152.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann (Ed.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (pp. 278-292). Stanford, CA: Stanford University Press.

Locke, E. A., and Latham, G. P. (1990). *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice Hall.

Lurie, N. H., and Swaminathan, J. M. (2008). Is timely information always better? The effect of feedback frequency on decision making. *Organizational Behavior and Human Decision Processes* 108(2): 315–329.

McCarty, P. A. (1986). Effects of feedback on the self-confidence of men and women. *Academy of Management Journal* 29(4): 840–847.

Moers, F. (2006). Performance measure properties and delegation. *The Accounting Review* 81(4), 897–924.

Northcraft, G. B., Schmidt, A. M., and Ashford, S. J. (2011). Feedback and the rationing of time and effort among competing tasks. *Journal of Applied Psychology* 96(5): 1076–1086.

Prendergast, C. (2002). The tenuous trade-off between risk and incentives. *The Journal of Political Economy* 110(5): 1071–1102.

Rabin, M. (2002). Inference by believers in the law of small numbers. *The Quarterly Journal of Economics* 117(3): 775–816.

Reichheld, F. F. (2003). The one number you need. *Harvard Business Review* 81(12): 46–54.

Reichheld, F. F. (2006). *The Ultimate Question: Driving Good Profits and True Growth*. Boston, MA: Harvard Business School Press.

Salmoni, A. W., Schmidt, R. A., and Walter, C. B. (1984). Knowledge of results and motor learning: a review and critical reappraisal. *Psychological Bulletin* 95(3): 355–386.

Savage, L. J. (1954). *The Foundations of Statistics*. New York, Wiley.

Schmidt, A. M., and Dolis, C. M. (2009). Something's got to give: The effects of dual-goal difficulty, goal progress, and expectancies on resource allocation. *Journal of Applied Psychology* 94(3): 678–691.

Sprinkle, G. B. (2000). The effect of incentive contracts on learning and performance. *The Accounting Review* 75(3), 299–326.

Sprinkle, G. B. (2003). Perspectives on experimental research in managerial accounting. *Accounting, Organizations and Society* 28(2–3), 287–318.

Tafkov, I. D. (2013). Private and public relative performance information under different compensation contracts. *The Accounting Review* 88(1), 327–350.

Taylor, S., and Thompson, S. (1982). Stalking the elusive vividness effect. *Psychological Review* 89(2): 155-181.

Thorndike, E. L. (1927). The law of effect. *American Journal of Psychology* 39(1/4): 212–222.

Tversky, A., and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin* 76(2): 105–110.

Van Buskirk, A. (2012). Disclosure frequency and information asymmetry. *Review of Quantitative Financial Analysis* 38(4): 411–440.

Figure 1—Experiment design

Treatment Group	Frequency of Feedback	Detail of Feedback
Group MA	Monthly	Aggregate
Group WA	Weekly	Aggregate
Group MD	Monthly	Detailed
Group WD	Weekly	Detailed

Figure 2—Timeline of the field experiment

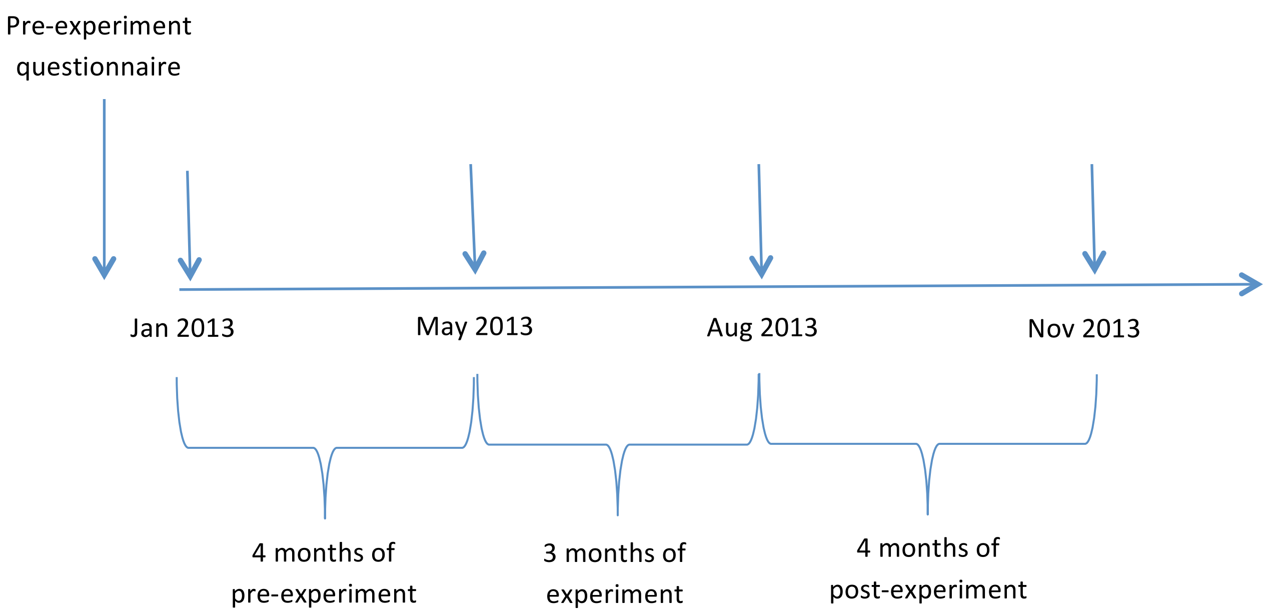

Table 1. Summary statistics

Panel A. Levels of Performance and Activity in the Pre-experiment and Experiment Period

	January - April 2013			
	MA	**MD**	**WA**	**WD**
Detractors	13.02%	14.23%	14.44%	14.53%
	(13.48%)	(15.73%)	(15.92%)	(16.22%)
No Detractors	30.25%	24.70%	25.71%	26.06%
	(46.01%)	(43.19%)	(43.73%)	(43.92%)
On Time	50.11%	53.76%	50.02%	48.55%
	(22.06%)	(21.41%)	(22.37%)	(21.35%)
PDA	76.50%	71.67%	76.39%	73.93%
	(21.39%)	(25.10%)	(22.37%)	(23.68%)
Number of services per month	49.17	60.79	62.07	53.00
	(39.87)	(50.13)	(69.41)	(47.83)
Number of surveys per month	15.21	19.48	19.93	17.50
	(13.79)	(18.09)	(27.34)	(20.36)
Number of professionals	90	92	273	265
	May - July 2013			
	MA	**MD**	**WA**	**WD**
Detractors	10.53%	8.37%	10.52%	11.15%
	(14.10%)	(12.22%)	(13.62%)	(15.05%)
No Detractors	39.10%	43.37%	38.05%	38.66%
	(48.89%)	(49.65%)	(48.58%)	(48.73%)
On Time	54.26%	55.76%	53.64%	50.79%
	(22.51%)	(22.06%)	(22.65%)	(23.43%)
PDA	79.30%	75.76%	77.40%	77.11%
	(19.24%)	(20.05%)	(20.00%)	(19.59%)
Number of services per month	44.34	51.49	56.43	46.57
	(37.72)	(41.06)	(65.92)	(42.71)
Number of surveys per month	12.39	15.07	16.78	14.52
	(10.99)	(13.41)	(24.69)	(18.21)
Number of professionals	98	104	294	291

Notes: This table reports the mean and standard deviation of the professionals' performance before and during the experiment. It also reports the mean and standard deviation of the number of services performed by the professionals and the number of services that were surveyed for customer satisfaction. Detractors measures the proportion of services performed by a given professional with a score of 6/10 or lower. No Detractors measures the proportion of observations with zero detractors in a month. On Time measures the proportion of services closed in on time. *PDA* is the share of services scheduled with the PDA.
Standard deviations in parentheses.

Table 1. Summary Statistics

Panel B. Pre-experiment Performance by Treatment Group

	Detractor	On Time	PDA
	(1)	(2)	(3)
MD	0.021	0.037	-0.051*
	(0.020)	(0.028)	(0.030)
WA	0.022	-0.002	-0.001
	(0.017)	(0.023)	(0.023)
WD	0.022	-0.016	-0.026
	(0.017)	(0.023)	(0.024)
Constant	0.091***	0.499***	0.763***
	(0.015)	(0.020)	(0.020)
p-value of F-test of joint statistical significance of interaction coefficients	0.615	0.156	0.173
Observations	2,589	2,703	2,599

Notes: This table shows Tobit regressions of a customer satisfaction metric and two measures of operational performance captured at the monthly level during the pre-experiment period: *Detractor* measures the proportion of services performed by a given professional with a score of 6/10 or lower; *On Time* measures the proportion of services closed in on time; *PDA* is the share of services scheduled with the PDA. *MD, WA*, and *WD* are treatment dummies that take a value of 1 or the professionals in the monthly-detailed, weekly-aggregate, and weekly-detailed treatments, respectively and 0 otherwise. Standard errors clustered by professional (individual) in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 2. The effects of feedback frequency and detail on customer satisfaction

	Detractor	Detractor	Detractor	Detractor	Detractor	Detractor	Detractor	Detractor	Promoter	Promoter	Score	Score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
MD	-0.022*	-0.022**	0.012		-0.033*	-0.034*	0.021	0.019	0.004		-0.040	
	(0.012)	(0.010)	(0.014)		(0.019)	(0.018)	(0.020)	(0.018)	(0.022)		(0.098)	
WA	-0.000	-0.005	0.014		0.002	-0.008	0.022	0.013	-0.003		-0.034	
	(0.010)	(0.009)	(0.012)		(0.017)	(0.015)	(0.017)	(0.016)	(0.019)		(0.082)	
WD	0.006	0.004	0.015		0.008	0.003	0.022	0.017	-0.009		-0.060	
	(0.011)	(0.009)	(0.012)		(0.017)	(0.016)	(0.017)	(0.015)	(0.019)		(0.082)	
Experiment			-0.025**	-0.054***			-0.041**	-0.104***	0.056***	0.088***	0.262***	0.463***
			(0.013)	(0.016)			(0.018)	(0.021)	(0.020)	(0.022)	(0.077)	(0.097)
Experiment * MD			-0.034**	-0.039**			-0.053**	-0.053**	0.025	0.032	0.141	0.185*
			(0.017)	(0.019)			(0.024)	(0.024)	(0.025)	(0.026)	(0.101)	(0.110)
Experiment * WA			-0.014	-0.015			-0.020	-0.018	-0.004	0.004	-0.017	0.014
			(0.015)	(0.016)			(0.021)	(0.021)	(0.022)	(0.023)	(0.089)	(0.095)
Experiment * WD			-0.009	-0.011			-0.014	-0.013	0.003	0.005	0.021	0.050
			(0.015)	(0.016)			(0.021)	(0.021)	(0.022)	(0.023)	(0.089)	(0.095)
Constant	0.105***	0.093***	0.130***	0.162***	0.047***	0.338***	0.090***	0.332***	0.541***	0.506***	8.216***	8.000***
	(0.009)	(0.009)	(0.010)	(0.007)	(0.015)	(0.007)	(0.016)	(0.011)	(0.017)	(0.008)	(0.071)	(0.046)
Estimation	OLS	OLS	OLS	OLS	Tobit	Tobit	Tobit	Tobit	OLS	OLS	OLS	OLS
Time effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Specialty effects	No	Yes	No	No	No	Yes	No	Yes	No	No	No	No
Individual effects	No	No	No	Yes	No	No	No	No	No	Yes	No	Yes
R-squared	0.004	0.076	0.018	0.313	-	-	-	-	0.019	0.343	0.021	0.342
Observations	2,133	2,133	4,722	4,722	2,133	2,133	4,722	4,722	4,722	4,722	4,722	4,722
p-value of F-test:												
Joint significance	0.025	0.034	0.172	0.178	0.067	0.099	0.104	0.099	0.437	0.513	0.240	0.243
MD=WD (Experiment)	0.006	0.009	0.116	0.135	0.020	0.031	0.072	0.061	0.429	0.401	0.255	0.272

Notes: This table analyzes the impact of the changes in feedback frequency and detail on various measures of customer satisfaction captured at the monthly level. *Detractor* measures the proportion of services performed by a given professional with a score of 6/10 or lower; *Promoter* is the share of services with a score of 9 or 10/10; and *Score* is the average score over all the services with a customer satisfaction survey in the period. *Experiment* is a dummy variable that takes value of 1 during the three months of the experiment and 0 otherwise. *MD, WA*, and *WD* are treatment dummies that take a value of 1 for the professionals in the monthly-detailed, weekly-aggregate, and weekly-detailed treatments respectively and 0 otherwise. Columns 1-4 and 9-12 use an OLS specification while columns 5-8 use a Tobit specification. The F-test for MD=WD (Experiment) in columns 1, 2, 5 and 6 compares the coefficients on the uninteracted dummy variables for the MD and WD treatments, in all other columns it compares the coefficients of the interactions of the MD and WD treatment dummies with the dummy for the experiment period.

Standard errors clustered by professional (individual) in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 3. The effects of feedback frequency and detail on operational performance

	On Time	PDA
	(3)	(6)
MD	0.037**	-0.044
	(0.017)	(0.029)
WA	0.012	-0.002
	(0.014)	(0.021)
WD	-0.006	-0.016
	(0.013)	(0.021)
Experiment	0.087***	0.055***
	(0.020)	(0.020)
Experiment * MD	-0.016	0.012
	(0.024)	(0.026)
Experiment * WA	-0.003	-0.017
	(0.020)	(0.020)
Experiment * WD	-0.021	-0.001
	(0.020)	(0.021)
Constant	0.206***	0.867***
	(0.044)	(0.010)
p-value of F-test of joint statistical significance of interaction coefficients	0.570	0.464
Time effects	Yes	Yes
Specialty effects	Yes	Yes
Observations	4,962	4,745

Notes: This table shows Tobit regressions of two measures of operational performance captured at the monthly level: *On Time* measures the proportion of services closed in on time; *PDA* is the share of services scheduled with the PDA. *Experiment* is a dummy variable that takes value of 1 during the three months of the experiment and 0 otherwise. *MD, WA*, and *WD* are treatment dummies that take a value of 1 or the professionals in the monthly-detailed, weekly-aggregate, and weekly-detailed treatments, respectively and 0 otherwise. Standard errors clustered by professional (individual) in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 4. Over-reaction to information vs dynamic incentives

	Detractor (1)	Detractor (2)	Detractor (3)	Detractor (4)
Dummy I	-0.149***	-0.255***	-0.255***	0.380**
	(0.048)	(0.060)	(0.060)	(0.181)
Dummy I * MD	-0.053	0.004	0.004	0.102
	(0.037)	(0.061)	(0.061)	(0.112)
Dummy I * WA	0.030	0.093*	0.093*	0.198**
	(0.031)	(0.054)	(0.054)	(0.099)
Dummy I * WD	0.045	0.091*	0.091*	0.200**
	(0.031)	(0.055)	(0.055)	(0.101)
Dummy II				0.440***
				(0.170)
Dummy II * MD				-0.046
				(0.092)
Dummy II * WA				0.012
				(0.078)
Dummy II * WD				-0.023
				(0.080)
Dummy III		-0.137***	-0.171***	-0.120**
		(0.049)	(0.061)	(0.053)
Dummy III * MD		-0.068*	0.057	-0.093*
		(0.041)	(0.066)	(0.049)
Dummy III * WA		0.014	0.111**	-0.005
		(0.033)	(0.054)	(0.040)
Dummy III * WD		0.034	0.127**	0.019
		(0.034)	(0.053)	(0.041)
Dummy IV			-0.123**	
			(0.051)	
Dummy IV * MD			-0.112**	
			(0.046)	
Dummy IV * WA			-0.022	
			(0.037)	
Dummy IV * WD			0.000	
			(0.038)	
Variable Definitions:				
Dummy I	Experiment	First week	First week	First week * Detr
Dummy II	-	-	-	First week * NoDetr
Dummy III	-	Later weeks	Later weeks * Detr	Later weeks
Dummy IV	-	-	Later weeks*NoDetr	-
F-test joint significance (p-value)				
Dummy I Interactions	0.007	0.082	0.082	0.137
Dummy II Interactions				0.871
Dummy III Interactions		0.017	0.076	0.039
Dummy IV Interactions			0.031	
Time effects	Yes	Yes	Yes	Yes
Specialty effects	Yes	Yes	Yes	Yes
Observations	17,372	17,372	17,372	14,454

Notes: This table shows Tobit regressions of a measure of customer satisfaction captured at the weekly level: *Detractor* measures the proportion of services performed by a given professional with a score of 6/10 or lower. *Experiment* is a dummy variable that takes value of 1 during the three months of the experiment and 0 otherwise. *MD, WA*, and *WD* are treatment dummies that take a value of 1 or the professionals in the monthly-detailed, weekly-aggregate, and weekly-detailed treatments respectively and 0 otherwise. *First Week* is a dummy variable that takes value of 1 for the first week of every experiment month and 0 otherwise, while *Later Weeks* takes a value of 1 for the other weeks of the experiment months and 0 otherwise. *Detr* is a dummy variable that takes value of 1 if the professional had at least one detractor in the first week of the month (column 3) or in the last week of the previous month (column 4) and 0 otherwise, and *NoDetr* takes value of (1-*Detr)*. We do not report the baseline coefficients on the *MD, WA*, and *WD* dummies for ease of presentation.
Standard errors clustered by professional (individual) in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5. Impact of prior period performance on performance variance

Performance last week of prior month		Frequency of feedback: Monthly	Weekly	W-test (p-value)
At least one detractor				
	Variance	0.0255	0.0533	10.232
	Observations	120	370	(0.001)
No detractor				
	Variance	0.0268	0.0376	2.114
	Observations	185	503	(0.146)
	W-test	0.313	14.822	
	p-value	(0.576)	(0.000)	

Note: This table presents the variance of performance in the first week of the month for professionals in the monthly and weekly feedback treatments respectively as a function of whether the last week of the previous month they had or not a detractor. Performance is measured as the number of detractors. W-test is the Levene’s (1960) robust test of difference in variances.

Table 6. Post-experiment effects

	Detractor (1)	Detractor (2)
Post Period I	-0.049**	-0.013
	(0.020)	(0.122)
Post Period I * MD	-0.022	0.064
	(0.024)	(0.092)
Post Period I * WA	-0.006	0.047
	(0.020)	(0.074)
Post Period I * WD	-0.008	0.043
	(0.020)	(0.075)
Post Period II		-0.068
		(0.116)
Post Period II * MD		-0.096
		(0.083)
Post Period II * WA		0.025
		(0.067)
Post Period II * WD		0.006
		(0.068)
Post Period III		-0.014
		(0.045)
Post Period III * MD		0.004
		(0.039)
Post Period III * WA		-0.004
		(0.032)
Post Period III * WD		0.009
		(0.033)
Variable Definitions:		
Post Period I	Post-Experiment	First Week Post * Detr
Post Period II	-	First Week Post * NoDetr
Post Period III	-	Later Weeks Post
F-test joint significance (p-value)		
Post Period I Interactions	0.788	0.911
Post Period II Interactions		0.370
Post Period III Interactions		0.943
Data	Monthly	Weekly
Time effects	Yes	Yes
Specialty effects	Yes	Yes
Observations	7,371	26,721

Notes: This table shows Tobit regressions of a measure of customer satisfaction: *Detractor* measures the proportion of services performed by a given professional with a grade of 6/10 or lower. Column 1 uses monthly data, and runs the same regression as in column 7 of table 2, with the additional variables corresponding to te post-experiment period shown here. Column 2 uses weekly data and runs the same regression as in column 4 of table 4, with the additional variables corresponding to the post-experiment period shown here. *Post-Experiment* is a dummy variable that takes value of 1 during the four months after the experiment (August to November) and 0 otherwise. *MD, WA*, and *WD* are treatment dummies that take a value of 1 or the professionals in the monthly-detailed, weekly-aggregate, and weekly-detailed treatments respectively and 0 otherwise. *First Week Post* is a dummy variable that takes value of 1 for the first week of every post-experiment month and 0 otherwise. *Later Weeks Post* takes a value of 0 for the first week of every post-experiment month and 1 otherwise. *Detr* is a dummy variable that takes value of 1 if the professional had at least one detractor in the last week of the previous month and 0 otherwise. *NoDetr* takes value of 1 if the professiona did not have any detractor in the last week of the previous month and 0 otherwise.
Standard errors clustered by professional (individual) in parentheses.
*** p<0.01, ** p<0.05, * p<0.1