

# An Experimental Examination of Assurer Reputation within a Multi-Assurance Service Context

Ananda Ganguly  
Assistant Professor

Joshua Herbold  
Doctoral Student

Mark E. Peecher  
Associate Professor

University of Illinois at Urbana-Champaign  
College of Commerce and Business Administration  
Department of Accountancy  
1206 South Sixth Street  
Champaign, Illinois 61820

August 2001  
version 1.2

Preliminary draft. Please do not cite without the authors' permission. We thank Jon Davis, Kathryn Kadous, Susan Krische, Joel Pike, and workshop participants at the University of Illinois at Urbana and University of Wisconsin at Madison for helpful comments on an earlier draft. We also thank several colleagues, especially Frank Nekrasz and Cathy Shakespeare, for help with subjects and in running the experiments. Comments will be appreciated.

## An Experimental Examination of Assurer Reputation within a Multi-Assurance Service Context

This paper examines the dynamics and economic implications of an assurance provider's (assurer's) reputation within a multi-service laboratory market setting. Herein, the assurer, by construction, has no incentive to defraud assurees and the assurer's reputation takes on a consumer research orientation, being defined as the assurance user's perception of the assurer's past performance and prospective future performance (Fombrun [1996]). We focus on both how an assurer's flagship-service reputation influences the initial reputation within two new assurance markets and how performance in these two new markets feeds back to affect the previously established flagship-service and overall reputation. The two markets differ in the degree to which the knowhow (i.e., knowledge and skills) required to perform within the new assurance market overlapped with knowhow required in the flagship market. We expected the difference in knowhow overlap to moderate market participants' evolving perceptions of the assurers' reputation.

Several interesting findings emerge. We observe both feedforward and feedback reputation effects. Flagship reputations feed forward to market participants' assessments of the assurer's new service reputation, and success and failure in new markets feed back to affect the assurer's flagship and overall reputation. Unexpectedly, the effects of success and failure do not reliably differ across the two new assurance markets. Finally, in contrast to conventional wisdom holding that reputation is hard to earn but easily lost, we do not observe a systematic link between assurer performance in new services and participants' bids for the assurer's attest report.

Collectively, our theory and findings have several potential implications. One, they affirm the theoretical importance of knowhow overlap for initial feeding forward of flagship reputations into new assurance markets, but they do not do so for subsequent feedback loops from the new market reputations to the flagship reputation. Two, they provide suggest that boundary conditions exist that constrain conventional wisdom holding that flagship reputations are easily damaged by failures in new, related markets. Our evidence appears more consistent with flagship encapsulation than with fragility. Three, because the dynamics of our three measures of reputation considerably differ, the use of caution is warranted when trying to map from studies that use attitudinal measures of reputation to studies that use economic measures of reputation.

**KEYWORDS:** reputation, assurance services, bids, probability assessment

## An Experimental Examination of Assurer Reputation within a Multi-Assurance Service Context

### *1. Introduction*

This paper examines the dynamics and economic implications of an assurance provider's (assurer's) reputation within a multi-service laboratory market setting. Herein, the assurer, by construction, has no incentive to defraud assurees and the assurer's reputation takes on a consumer research orientation, being defined as the assurance user's perception of the assurer's past performance and future performance prospects (Fombrun [1996]). We focus on both how an assurer's flagship-service reputation influences the initial reputation within two new assurance markets and how performance in these two new markets feeds back to affect the previously established flagship-service and overall reputation. We expect these reputation dynamics to be moderated by market participants' initial and evolving perceptions of the inter-service transferability of assurers' knowhow (i.e., knowledge and skills, Teece [1980]).

Improving our understanding of reputation dynamics within assurance settings is important because CPA and non-CPA assurers are expanding the range of assurance services they offer. Some assurers seemingly have tried to lever reputations within a flagship market to help penetrate related new markets. Underwriter's Laboratory has expanded its core of product safety testing services to include a variety of new verification and inspection services (Underwriters Laboratory, Inc. [2000]), and many CPA professional service organizations have expanded the scope of assurance services offered, attempting to lever an established reputation in financial-statement auditing (AICPA [1999]; Boritz and Cockburn [1998]).

While CPA assurers likely can lever their financial-statement auditing reputations to penetrate some new assurance markets (AICPA [1997]; Elliot [1997, 1998]; Watts and

Zimmerman [1983]), predicting the extent and ramifications of success or failure within specific new markets is difficult. Difficulty arises because we possess a limited theoretical understanding of the determinants of reputation dynamics within a multi-service and multi-period context and because we know little about how perceptions of assurer reputation map into economic behaviors such as actually paying assurers to render services. Attempting to lever a flagship service reputation also is risky. Conventional wisdom holds that reputation can take years to build, but that it can be destroyed in little time (e.g., Fredeen [2000]). Similarly, CPAs' entry into new markets, especially ones that feature unreliable or unfamiliar assertions, could elevate the risk of poor new-market performance which, in turn, could erode CPAs' established reputations (Sundem, Dukes, and Elliott [1996]).

The organizational economics and consumer behavior literatures identify market participants' perceptions about the overlap in the knowhow required to perform a flagship service and a new service as a likely determinant of reputation feedforward and feedback loops (e.g., Teece [1980]; John, Loken & Joiner [1998]). This overlap influences the extent to which market participants will expect an assurer's track record in a flagship service to replicate in a new service. As an example, CPAs likely can transfer their financial-statement auditing reputation more easily to attestations of mutual fund performance than to attestations of health care performance statistics. Indeed, we hypothesize that greater perceived knowhow overlap leads to greater initial reputation transfer. We also, however, hypothesize that greater perceived knowhow overlap leads to greater downstream risk to the assurer's flagship and overall reputation should poor performance be realized in the new market. That is, we hypothesize that failure in newer services with higher, as opposed to lower, perceived knowhow overlap will do more damage to the assurer's flagship and overall reputation. Finally, we hypothesize that ongoing realizations of successful and unsuccessful

new-service performance will affect participants' assessments of the likelihood that assurers will be successful in attestation, and their bids for assurers' attest reports.

To test our hypotheses, we conducted a series of laboratory markets. In each market, sellers sold buyers one of three hypothetical commodities -- gold, silver, or diamonds -- during a series of trading periods. Commodity quality was high or low, observable *ex ante* by sellers (not by buyers), and randomly determined (i.e., with a fixed probability of high quality equal to 50%). High quality assets had a higher resale value for the buyers, and they could observe quality *ex post* (after trading within a period). In our experimental settings, sellers could not signal quality credibly, since they were not liable for asking high prices when product quality was low. Thus, third-party attestation was potentially valuable to buyers (Dopuch & King, 1991).

Before each period, buyers submitted bids for the right to privately view the attest report. Only the buyer with the highest bidder privately viewed the attest report prior to trading. After each period, the assurer's success or failure was made public by oral statement and by posting it on a large running assurer scoreboard. The assurer's flagship reputation was formed by having only one commodity traded (gold) for the first five periods. The assurer successfully attested to the quality of the commodity during each period.

Entry into the first new assurance market occurred after the first five periods. The between-subject experimental manipulation was whether the first new market featured "higher" knowhow overlap (i.e., silver) or "lower" knowhow overlap (i.e., diamond). After a single new-market period, a shift back to the flagship occurred, followed by two more periods within the first new assurance market. Thereafter, the nine period sequence repeated except it featured a different new assurance market. Repeating the sequence in this manner yielded a within-subjects experimental manipulation of the ordering of entry into the two

new assurance service markets (higher overlap first or lower knowhow overlap first). The sequences appear in exhibit 1, panel C.

Consumers' perception of the assurer's reputation was elicited in three ways. We directly solicited buyers' perceptions of the assurer's reputation (overall, and separately for flagship and the two new services) at the end of every three periods (hereafter, periodic assessments). Additionally, before each period, we solicited buyers' likelihood judgments as to whether the assurer would succeed or fail in that period (hereafter, probability estimates), as well as the amount (in a fictional currency) that they were willing to bid to privately obtain the assurer's report prior to trading (hereafter, bids). To increase the reliability of bid data, we held a second-price, sealed-bid auction. The attest report gave the assurer's opinion of asset quality, and the probability of an accurate attest report varied depending on whether the service was flagship or new (and differed across the two new services to reflect differential knowhow overlap -- we explain the mechanism used to accomplish this later).

Our main findings concern both feedforward effects from the flagship service reputation to reputations for the two new services and feedback effects from reputations of the two new services back to the flagship (and overall) reputations. In terms of the feedforward effects, we observe some evidence that the initial transfer of a flagship reputation to new service reputation is significantly greater when the perceived inter-service overlap in knowhow is greater. When a greater extent of initial reputation transfer occurs, market participants provide higher initial probability assessments (or report higher reputation opinions suggesting) that the assurer will succeed in the new assurance service.

In terms of the feedback from the new services to the flagship service, the success or failure of assurers in new markets significantly affects both the overall and flagship service reputation in the very short term. In the medium to longer terms, the overall reputation

continues to mirror performance in all service sectors, but the flagship service reputation becomes encapsulated (i.e., protected from being substantially damaged, explained later in theory development). Unexpectedly, however, the effects of success and failure are not significantly different in the very short term across the two new services even though market participants judge one new service to have greater knowhow overlap with the flagship service. Finally, although negative effects on the assurer's overall and flagship reputations measured by periodic assessments arise as a result of the assurer's ongoing degree of success in the new services and persist throughout all market periods, few negative effects are evident in either participants' probability estimates of assurer success or bids for the assurer's attest report in the flagship service.

Collectively, our theory and findings have several potential implications. One, they affirm the theoretical importance of knowhow overlap for initial feeding forward of flagship reputations into new assurance markets, but they do not do so for subsequent feedback loops from the new markets to the flagship reputation. Two, they provide suggest that boundary conditions exist that constrain conventional wisdom holding that flagship reputations are easily damaged by failures in new, related markets. Our evidence appears more consistent with flagship encapsulation than with fragility. Three, because the dynamics of our three measures of reputation considerably differ, the use of caution is warranted when trying to map from studies that use attitudinal measures of reputation to studies that use economic measures of reputation.

## *2. Theoretical Development and Hypotheses*

Given the immense size of the entire assurance-service market, CPA assurers naturally would like to extend their menu of assurance services, but without damaging their hard-earned reputations (e.g., King and Schwartz 1998). The services currently provided by

CPA assurers generate considerable knowledge of client organizations. Both intuition and observation of the success of accounting firms suggest that economies of scope exist (AICPA [1997]). To the extent that market participants believe that the 'knowhow' within assurance firms is transferable to new services, the firms can attempt service extensions and reap further profits from economic scope (Teece [1980]). Market participants' perceptions of the transferability of CPAs' knowhow, not regulatory mandate, will constrain CPAs' competitive position in the future assurance service marketplace. As Elliot (1997) points out, there is no statutory franchise for CPAs in assurance services outside of financial statement audits. If CPAs do not use their economies of scope and lever their knowhow within new assurance services, then competitors will, and CPAs will lose a potentially lucrative revenue source.

Despite the potential lucrativeness, some argue that CPA assurers should exercise considerable caution before leveraging their established reputations to gain access to some new assurance markets (Sundem, Dukes, and Elliott [1996]). Clearly, there exists the danger of negative feedback if perceived poor performance obtains in the new market. Anecdotal and empirical evidence demonstrate that mistakes within the financial statement audit arena can hurt an audit firm's reputation. For example, inspection reports of specific companies by the U.K. Department of Trade are sometimes critical of the work done by the company's auditors. Firth (1990) finds that such critical reports are associated with a small negative abnormal return for the auditor's other publicly traded clients (other than the client under investigation) and with slower audit firm growth rates (also see Wilson and Grimlund [1990]).

The initial and downstream net benefits associated with leveraging a flagship-service reputation depend in part on reputation dynamics between flagship and new services.

Consumer research shows that consumers evaluate brand extensions based on the "fit" between the overall brand and an extension (e.g., Aaker and Keller [1990]; John, Loken, and Joiner [1998]). Fit can be operationalized as the perceived similarity between the attributes of a flagship product and new product extensions (Ahluwalia and Gürhan-Canli [2000]). In assurance service markets, a salient attribute that likely affects the perceived similarity or "fit" is perceptions about the inter-service overlap in knowhow (i.e., knowledge and skills). We expect that when market participants perceive a greater level of knowhow overlap between a flagship and new assurance service, the fit increases. As the fit increases, the potential for initial reputation transfer (from the flagship to the new service) is heightened. When greater reputation transfer from a flagship to a new service occurs, market participants may be willing to pay more for the new assurance service.

However, greater initial reputation transfer also leads to an increased risk of (negative) reputation feedback (from the new to the flagship service) in the case of poor performance in the new assurance service. Market participants' mental representations of the assurer's overall and service-specific reputations likely are associated in memory (Keller [1993]).<sup>1</sup> The memory associations across different service reputations likely increase in strength with knowhow overlap (Collins and Loftus [1975]). The stronger the associations across mental representations of flagship reputations, new-service reputations, and overall reputations, the more likely that performance in a new service will affect the mental representations of that one service's reputation and both the reputation for other services as well as overall reputation. And, aside from memory effects, to the extent to which there is an inter-service overlap in knowhow, market participants purposely will treat success and failure as a signal about future flagship performance.

---

<sup>1</sup> Most memory models assume associative links between knowledge and concepts (Wyer and Srull [1989]).

Research in the consumer behavior area also has demonstrated that a product's reputation consists of both an overall brand component and a product-specific component. Reputation spillovers occur because information that affects the product specific component of a particular product's reputation will also cause a change in the overall brand reputation, which then affects the reputation of all products under that brand umbrella (Aaker and Keller [1990]). Loken and John (1993) find that negative aspects of a brand extension can dilute (damage) an overall brand reputation when the original and extension products have differing levels of quality. They also find, however, that negative effects to flagship services are less pronounced when consumers perceive there to be little 'fit' between the original and new products. If a sufficiently small degree of fit exists, a flagship service could be 'encapsulated' and thus escape damage even when an assurer's overall reputation suffers as a result of poor performance in new services (John, Loken, and Joiner [1998]).

Integrating the organizational economics and consumer behavior research, we hypothesize that: (1) Greater perceived knowhow overlap leads to greater initial reputation transfer, but also carries greater risk of damage to the overall and flagship reputations in the event of poor performance in the new service market; (2) Changes in reputations due to initial transfers and the negative (or positive) feedback effects will influence market participants' likelihood judgments that assurers will perform well in specific engagements for flagship and new service markets; and, (3) Such changes in reputation also will influence how much they bid for assurer reports in specific engagements for flagship and new service markets. Collectively, these three propositions imply that reputation spillovers can significantly affect an assurer's stream of economic rents.

An important limitation of the consumer behavior research on reputation dynamics within brand extension contexts is the reliance on a one-shot experimental setting instead of

within a repeated-interaction experimental setting (Hertwig and Ortmann, 2001). Reputation dynamics likely subsume these "one-shot" reactions whenever, over time, reasonably separable track records for flagship and newer services emerge.<sup>2</sup> Whether a flagship initially is "artificially" diluted by early, poor new market performance becomes less critical if market participants restore the flagship reputation after learning more about the association between flagship and new service performance. Consequently, the possibility that flagship reputations can be reasonably encapsulated is greater over the longer term than over the shorter term, provided that market participants learn. A degree of learning likely occurs when market participants receive sufficiently clear and timely feedback about the assurer's performance across different markets (Einhorn [1980]). Learning from outcome feedback over repeated interactions occurs when participants have an evolving causal theory of the domain prior to initial interactions (Camerer [1981]).

A second limitation of such consumer behavior research is that studies typically secure attitudinal measures of brand image such as experimental participants' beliefs about product and brands (i.e., favorableness or reputation of an overall brand) without linking these attitudinal measures into a differential willingness to buy products under a brand umbrella, including flagship and "extension" products.

In our study, we attempt to overcome these two limitations by including repeated interactions and by observing the relation between an assurer's overall and service-specific reputations and market participants' willingness to bid for the assurer's services. While assurer reputation likely influences market participants' willingness to bid for an assurer's services, a number of other factors also likely affect their bids. By using an approach that makes attitudinal measures of reputation especially salient to market participants, the

---

<sup>2</sup> For extreme cases of initial damage to a flagship or an overall reputation, an assurer could immediately

opportunity to observe a link between such measures and expected assurer performance and bids is strong.<sup>3</sup> We next discuss our measure of reputation and its relationship to two less direct measures of reputation.

Our primary measure of reputation is provided by sets of subjective periodic assessments completed by subjects after every third period (see Exhibit 1). Each set of periodic assessments includes a measure of the overall assurer reputation, the assurer's flagship-service reputation, and the assurer's reputation in two newer assurance service markets. The use of periodic assessments or surveys is common in the consumer literature and brand equity research. For example, in researching the effect of brand-specific associations on brand extensions, Broniarczyk and Alba (1994) asked subjects to evaluate hypothetical brand extensions of actual (existing) brands. Loken and John (1993) used a similar dependent measure to study whether brand extensions can dilute brand beliefs, and Dacin and Smith (1994) surveyed subjects to demonstrate the importance of brand portfolio characteristics to the evaluations of brand extensions (see Shocker, Srivastava, and Rueker (1994) for a review of these and other papers). Finally, in a study on the reputation of strategic groups, Ferguson, Deephouse, and Ferguson (2000) used the ratings given to insurance companies by A. M. Best Company, Standard & Poor's Corp., and Weiss Ratings, Inc.

Since we believe that it is important to demonstrate that reputation is economically meaningful, we link this primary measure of reputation to two indirect measures of

---

withdraw from the new assurance service market.

<sup>3</sup> In natural environments, it is unclear whether the relation between assurer reputations and market participants' willingness to pay for assurer services would be strong enough for systematic relations to obtain. One cannot unambiguously make the link to "reputation" based on archival research. Archival research typically does not collect attitudinal measures of reputation so as to allow one to infer that, for example, a decline in the price for an assurer's service stems from declines in assurer reputation. Declines in prices could reflect a number of alternatives including declines in the perceived value of receiving assurance due to things

reputation. Specifically, we solicited market participants' probability assessments that the assurer's report would be correct as well as their bids for the attest report. We next separately discuss these two measures, which were collected before trading in every market period.

A number of studies operationalize assurer reputation as the perceived probability that the assurer will exert high effort, or the perceived probability that the assurer is of a given type (where assurer type determines the assurer's accuracy, effort, or honesty).<sup>4</sup> Over repeated interactions, these judgments likely depend on subjects' perceptions of the assurer's ability as well as on subjects' perceptions about the time series properties of assurer performance. These time-series properties could include participant-specific hypotheses about whether an assurer's success or failure "streak" can or cannot last and the like (e.g., gambler's fallacy and other illusion of control phenomena). Another noteworthy aspect about such probabilities is that they do not measure the subjective value of the report; they only measure one component of the subjective value.

In contrast to such probabilities, participants' second-price, sealed-bids for the right of private, pre-trading access to the report is a measure of the report's subjective value. Thus, logically, higher assurer service-specific reputation should increase the amount one is willing

---

such as high assertion reliability, low correlation between improved knowledge about assertions and assuree's own profits, and so on.

<sup>4</sup> Prior research using probabilities as the dependent measure include Datar and Alles (1999), where the auditor's local reputation with the manager is the manager's belief about the auditor's effort level (i.e.,  $P[\text{Auditor exerts high effort}]$ ). In Mayhew (2001), the price for audit services was fixed, and the manager made a choice regarding which auditor to hire. The value of the report was held constant, so the manager's decision was based on her perceived probability of the auditor exerting high effort. Similarly, in Charness and Goupa (2000), subjects could choose to buy information about an asset for a fixed price (this decision was based upon the subjects' assessments of the likelihood that the information would be accurate), and in Yaniv and Kleinberger (2000), the authors tested the effect of advisor reputation on advisees' willingness to buy advice (the price and the value of the advice were held constant, so the decision was based on subjects' assessments of the probability that the advice would be accurate).

to bid.<sup>5</sup> If one believes the assurer always is right, the ceiling for bids is the subjective value of perfect information on commodity quality within the market setting. Of course even second-price, sealed-bids potentially bring unwanted strategic elements into play (Kagel [1995]). Some subjects, for example, may value simply winning the report, and thus bid to win the report rather than bidding their subjective "informational" value of the report.<sup>6</sup> In addition, subjects' judgments about the properties of the underlying time series of commodity quality may influence the value of having access to the attest report. If a subject believes they know the trend in quality (i.e., gambler's fallacy and other illusion of control phenomena could be operative), there would be little expected informational value in bidding for the assurer's attest report.<sup>7</sup>

### 3. Methodology

A general outline of the experiment is shown in Exhibit 1, panel A. Subjects were 317 undergraduate accounting students at a large university, who participated in one of twenty-two experimental market sessions for a nominal amount of course credit. In addition, subjects competed (in the markets) for profits denominated in francs (a fictional currency). Each subject's relative (normalized) number of francs determined the number of lottery tickets they won, with one ticket drawn per session for a \$50 prize.

---

<sup>5</sup> Prior research that included bids (or a related measurement) as a dependent measure includes King (1996), who used the difference between either buyers' forecasts or sellers' reports and actual dividends to determine information gap costs, and Bandyopadhyay and Kao (1999), who claim that audit fee premiums for Big 6 firms possibly reflect brand name reputation rather than monopoly or oligopoly rents.

<sup>6</sup> In future versions of this paper we may examine at least one further measure of assurer reputation related to the value of the attest report. Logically, we expect that the following relation should hold: Probability  $\times$  Value = Bid. Thus, the "value" of the report can be measured by Bid/Probability, or as the incremental bid for each  $x\%$  increase in probability. A large number of studies (mostly analytical in nature) have used this measure of value (or something similar) to investigate reputation. Camerer and Weigelt (1988) is representative of these studies, and in it they examine the joint effects on reputation of two parameters: payoffs and the probability of a player's type. Other papers based on this framework include Fudenberg and Kreps (1987), Neral and Ochs (1992), Jung, Kagel, and Levin (1994), Jackson and Kalai (1999), and Park (1999). Ganguly, Kagel and Moser (1994, 2000) show that subjects appear to bid their expected payoffs based on their subjective (sometimes biased) probabilities and known payoffs.

During the class period prior to the experimental markets session, subjects were introduced to the computer program used for trading in the markets and participated in a practice second-price, sealed-bid auction. During the same class period, subjects also completed a three-section “quiz.” Each section of the quiz related to one of three types of assets traded in the market sessions. The first (Analytical) section of the quiz contained questions similar to Raven matrices that measured analytical (specifically, “relational”) reasoning ability.<sup>8</sup> The second (Financial-Analytical) section of the quiz was designed to mimic the first section’s testing of “relational” reasoning in an accounting setting (i.e., the questions in the second section had to do with accounting patterns, ratios, and relations). The third section of the quiz (Financial-Recall) consisted of questions that required the correct recall of accounting facts (i.e., accounting rules and definitions).

The three quiz sections operationalized the different types of knowhow (i.e., knowledge and skills) that would determine the accuracy of the attest report for each of three types of assets (gold, silver, or diamonds) to be traded in the later experimental market sessions.<sup>9</sup> Specifically, the hypothetical assurer’s accuracy with respect to the quality of gold was related to the Analytical section score of three of the top performers on the quiz. Similarly, the assurer’s accuracy rate for silver (diamonds) was related to the Financial-Analytical (Financial-Recall) score of three of the top performers in the quiz. Thus, the cognitive processes required for determining the quality of gold and silver were designed to

---

<sup>7</sup> Although we emphasized to participants that the trend in quality was completely random, it is unlikely that participants understood the implications of randomness (see Bloomfield and Hales [2001] for a recent discussion).

<sup>8</sup> The Raven Progressive Matrices Test (Raven, Raven & Court [1995]) was developed to be a measure of analytical intelligence. *Analytical intelligence* is generally defined as “the ability to reason and solve problems involving new information, without relying extensively on an explicit base of declarative knowledge derived from either schooling or previous experience.” (Carpenter, Just, and Shell [1990]) This form of intelligence has also been labeled *fluid intelligence* in the psychology literature (see, e.g., Cattell [1963]).

<sup>9</sup> We used a hypothetical assurer such that assurer's accuracy rate was related to the accuracy rate of top performers on the quiz. This operationalized differential overlaps in the knowhow required to attest to the quality of specific assets.

be internalized by participants to be similar along one dimension (both required analytical ability), while the cognitive processes required for determining quality of silver and diamonds were similar along another dimension (both required knowledge of financial accounting). The cognitive processes required between gold and diamonds were dissimilar along both dimensions (Analytical vs. Financial-Recall). The relationship between the three sections of the quiz and the assurer's accuracy rate in the three types of assets is shown in Exhibit 1, panel B. Students were informed about all the above facts, except that the content-type differences in the three parts of the quiz were not explicitly explained to them.

During the experimental market sessions (which followed a few days after the practice second-price auction and quiz described above), subjects bought the three types of assets from two sellers who were confederates of the experimenters. Each market session consisted of several periods. The order of the periods is given in Exhibit 1, panel C. All market sessions consisted of at least the first 17 periods indicated in Exhibit 1, panel C. In some other sessions, time constraints prevented us from completing additional periods. For each market period, the confederate sellers worked from a schedule of increasing asset prices (i.e., the marginal price was increasing), and continued to sell assets to the highest bidder as long as the bid price exceeded the marginal price shown on the schedule. Subjects could buy up to a maximum of five units of the asset during any given period. During each market period, subjects bought only one of the three asset types. Whether the asset being sold in a period was gold, silver or diamonds was announced in advance. Subjects were also told that the asset being traded in a given period could be of either high or low quality (e.g., real gold or fool's gold), with equal prior probability. All assets traded in a given period were of the same quality. Subjects earned profits by re-selling the assets to a hypothetical buyer (at pre-

determined values) at the end of each period (high quality assets had a higher resale value than did low quality assets).

Although subjects did not know in advance the true quality of the asset being traded in a period, they did know the prior probability of high quality, which was 50%. In addition, they had the opportunity to buy a report from an assurer on the quality of the asset. The assurer's report for each period was sold to the highest bidder in a second-price, sealed bid auction. (The theoretical dominant strategy in private value auctions conducted in this way is to bid one's true willingness to pay. First-price auctions may be affected by a subjects' risk preferences, but second-price sealed-bid auctions should theoretically be free of this problem (Vickrey [1961]). As explained above, the accuracy of the assurer's report in each period was related to top scores on the quiz that students previously had completed. This accuracy rate was operationalized in each market period by having one of the subjects randomly select (with replacement) from a set of cards labeled "success" or "failure" (the subject selecting the card could not see which type of card she selected). The number of cards of each type corresponded to the top three quiz scores.

After the assurer's report was sold in each period to one of the market participants, the market for gold, silver, or diamonds opened in that period. In these markets, assets were sold to the highest bidder in a double auction conducted on MUDA software (Plott and Grey [1990]). At the end of each period, the true quality of the asset and the report that the assurer had made were both announced and written on a running scoreboard at the front of the room.<sup>10</sup> Subjects then calculated their profits for the period, and the process was

---

<sup>10</sup> Some may point out that for financial-statement auditing and other "credence" services true quality remains elusive. Without prejudice to our design or procedures one could assume that "success" means only that the assurer correctly judges that an asset can be passed from the lab market buyers to the next round of buyers without any revelation of a quality difference. We do not have to assume our assurer determined truth, per se.

repeated for the next period. After all periods had been completed, subjects filled out a questionnaire containing demographic information and various manipulation checks.

As described earlier, the asset being traded in a given period was either gold, silver, or diamonds. The sequence of assets traded is shown in Panel C of Exhibit 1. To establish the “flagship service” reputation (attesting to the quality of gold), gold was traded in periods 00 through 04 of each market session and the assurer always was successful in identifying asset quality.<sup>11</sup> In approximately half of the sessions, the first *new* asset that subjects saw (i.e., in Period 05) was silver, and in the other half it was diamonds. After trading the new asset in periods 05, 07 and 08, the sequence was repeated using the other new asset (i.e., the one that subjects had previously not seen). In other words, the sequence shown in Exhibit 1, panel C for periods 00-08 was repeated starting with period 09, replacing the silver (diamonds) traded in periods 05, 07, and 08 with diamonds (silver). Subjects’ probability assessments of the assurer’s accuracy rate were collected at the start of each period in each market session. In addition, starting with period 02 and for every third period thereafter, subjects made a periodic assessment of the reputation of the assurer. The periodic assessments required the subjects to rate the assurer on overall reputation, as well as separately for reputation in the gold, silver and diamond markets.

#### 4. *Experimental Findings*

Manipulation Checks. After the in-class quiz, participants were asked two follow-up questions. Each question told participants that a hypothetical Joe Doaks answered 4 questions correctly from a named section of the quiz (e.g., analytical section) and that he also

---

<sup>11</sup> We use five periods for reputation building purposes whereas Camerer and Weigelt (1988) found that six periods was sufficient for reputation building. Our subjects, however, were provided with a concrete basis for forming a prior that the assurer would perform well (i.e., reflect performance of a “top performer” on the aforementioned in-class quiz). Nevertheless, in choosing 5 periods instead of 6, we may have erred on the side of having a relatively fragile “flagship.” However, a relatively fragile flagship simply would give feedback loops in the new assets markets a greater chance of affecting the flagship reputation.

correctly answered 2 questions and 4 questions on the two remaining sections of the quiz, without naming which sections. It then asked participants whether it was more, less, or equally likely that Joe Doaks correctly answered 4 questions on one named section of the quiz (e.g., financial-analytical) as opposed the other named section (e.g., financial recall). The first question thus addressed whether the participants believed that the knowhow required to answer the abstract analytical questions overlapped more with the knowhow required to answer questions from one or the other remaining quiz sections. The second question addressed whether participants believed the knowhow required to answer financial-recall questions overlapped more with knowhow required for either the financial analytical or the analytical questions.

A continuous, five-point scale was used for both questions (much more likely 4 correct on "X", somewhat more likely 4 correct on "X", equally likely 4 correct on "X" or "Y", somewhat more likely 4 correct on "Y", and much more likely 4 correct on "Y"). For the first question, ratings below the midpoint indicate a greater perceived knowhow overlap between analytical and financial-analytical than between analytical and financial-recall. The mean rating of 2.384 is significantly lower than the midpoint of 3 ( $|t|=12.07$ ,  $p<0.001$ ). For the second question, a rating above the midpoint indicates a greater perceived knowhow overlap between financial-recall and financial-analytical than between financial-recall and analytical. The mean rating of 3.308 is significantly greater than 3 ( $|t|=5.68$ ,  $p<0.001$ ). These findings suggest that participants understood the intended knowhow overlap between the three quiz sections.

Two additional manipulation checks address whether participants mapped from the quiz to the assets actually traded in the market sessions. First, we simply examined whether participants' answers to a matching exercise in a post-market questionnaire indicated that

they made the link. Approximately 95%, 80%, and 81% of participants matched the analytical questions to gold, the financial-analytical questions to silver, and the financial-recall questions to diamonds, respectively. Second, we examined whether variation in participants' periodic assessments of the assurer's silver reputation explained more of the variation in the assurer's overall reputation than did variation in participants' periodic assessments of the assurer's diamond reputation. The idea underlying this check is that while overall reputation likely reflects all three service reputations to some degree, the greater knowhow overlap between silver and gold makes the assurer's silver reputation more relevant than the diamond reputation to the overall reputation. As shown in Table 1, the assurer's silver reputation consistently explains more variation in the assurer's overall reputation than does the assurer's diamond reputation. Averaging across all the sets of periodic assessments, the silver reputation explains 27% of the variation in overall reputation while the diamond reputation explains only an average of 11% of the variation in overall reputation. Further, the average incremental increase in the adjusted R-squared by adding diamond (silver) reputation into models that include the silver (diamond) reputation is 0.83 (16.67). From these data, we conclude that participants understood the intended manipulation of knowhow overlap useful for attesting to gold, silver, and diamonds.

Overall Analysis of All 18 Periods. Previous research suggests that subjects base their buy/sell decisions primarily on their (biased) subjective probability estimates of high or low outcomes (see, e.g., Ganguly, Kagel and Moser [1994, 2000]; Ackert, Church and Zhang [2001]). As a benchmark, it is thus reasonable to expect bids and probabilities to be positively correlated in our experiment (see footnote 6). Indeed, we find that between subjects, probabilities are significantly correlated with bids in 13 of 18 periods ( $0.10 < r <$

0.24;  $p < 0.01$  in 9 periods;  $p < 0.05$  in 2 periods; and  $p < 0.10$  in 2 periods).<sup>12</sup> Periods where correlations were not significant included periods 04, 12, and 13, where subjects had previously experienced a long run with the flagship asset (gold), and may have been exploring new bid strategies that were unrelated to their probability judgments; and periods 06 and 07, where subjects may have experienced some degree of uncertainty in their probability judgments after experiencing a new asset for the first time (see Exhibit 1 for order of assets traded). For the most part, the bid-probability correlation was positive and significant at inflexion points in the experiment, where subjects judged the implications of a shift in the asset traded. That is, the data indicate that the changes from one asset to another seem to make subjects pay more attention to their probability judgments when determining their bids; however, when the stimulus remains constant (i.e., long runs of gold), then other considerations seem to come into play.

Within subjects, the bid-probability correlation is also positive and significant ( $r = 0.15$ ,  $p < 0.001$ ). In addition, the correlation is stronger in the second half of the experimental sessions (for periods 09-17,  $r = 0.20$ ,  $p < 0.001$ ) than in the first half (for periods 00-08,  $r = 0.13$ ,  $p < 0.001$ ). Thus it appears that the subjects' use of non-probability information in decision models decreases with experience in our markets. To the extent that the use of non-probability information may be due to cognitive biases, this finding is consistent with previous research documenting that subjects 'learned to be unbiased' over repeated trials with feedback in certain circumstances (Ganguly, Kagel and Moser [1994, 2000]; Ackert, Church and Zhang [2001]).

Between subjects, probabilities are also strongly positively correlated with the periodic assessments. Recall that periodic assessments were collected at the end of every 3<sup>rd</sup>

---

<sup>12</sup> These p values are unadjusted. Applying the Bonferroni adjustment reduces the significance to periods 5 and

period (see italicized in Exhibit 1, Panel C) and the probability estimates were collected at the beginning of every period. In fact, each periodic assessment was collected before a period in which gold was traded. The flagship (gold) periodic assessment of reputation was thus highly positively correlated with the immediately following probability estimate of accuracy in the flagship (gold) market ( $0.37 < r < 0.67$ ;  $p < 0.01$  in all cases, even with Bonferroni adjustment).<sup>13</sup> The periodic assessments of silver and diamond attestation are also highly correlated with immediately following periods in which these assets were traded. Thus, dividing the data set into two groups based on the sequence (i.e., based on which asset was presented in period 05, see Exhibit 1 Panel C), the periodic assessments for silver/diamond collected in periods 2, 5, 11 and 14 are highly correlated with the probability estimates for attester success in periods 5, 7, 14 and 16 respectively ( $p < 0.01$  in each case). Within subjects too, the correlation of periodic assessments and probability estimates is positive and significant (e.g.,  $r(\text{PA-overall, probability}) = 0.24$ ,  $p < 0.001$ ;  $r(\text{PA-gold, probability}) = 0.47$ ,  $p < 0.001$ ).

The close relationship of periodic assessments of gold in periods 2, 5, 8, 11, 14 and 17 with probability estimates in (gold) periods 3, 6, 9, 12, 15 and 18 do not hold for bids in the latter set of periods. Periodic assessments of gold reputation and immediately succeeding gold bidding are not correlated at conventional significance levels in any period. Thus, probability estimates are better predictors of bidding behavior, especially over time, than periodic assessments. Periodic assessments seem to be more “sticky” and represent a “pure” measure of the consumer behavior notion of reputation. Probability estimates, in comparison, are more “mercurial” and may be affected by other effects, such as gamblers’

---

<sup>13</sup> only, with  $r=0.23$  and  $r=0.24$  respectively.

fallacy-type biases (i.e. expecting an outcome that has not occurred in a long time, or expecting a “run” of recent good luck to continue in the near future; see fn. 7). Besides, as pointed out earlier, bids themselves correlate with probability estimates only when a new or atypical asset is experienced. It is noteworthy that while periodic assessments do not directly correlate with bids, there do seem to be economic returns to building reputation *in the long run* (periodic assessments and probability estimates are positively correlated, and probability estimates are correlated with bids in asset-transition periods). However, issues such as immediate prior assurer success or failure seem to also be important factors in predicting *short-run* economic return. We next separately examine the behaviors of bids, probabilities and periodic assessments of reputation.

Analysis of Bids for All 18 Periods. Figure 1 presents the median bids over all 18 periods, separated by the new asset sequence.<sup>14</sup> Ocular analysis of the data indicates that first-time experiences with a new similar asset (silver, comprised of periods 5, 7, and 8 for asset\_05=2 and periods 14, 16, and 17 for asset\_05=3) causes bids to remain, on average, consistent with prior levels in the short run; but experiencing a new dissimilar asset (diamonds, comprised of periods 5, 7, and 8 for asset\_05=3 and periods 14, 16 and 17 for asset\_05=2) makes bids fall in the short run. This is true regardless of whether the similar or dissimilar asset is experienced first.

Across assets, the first new-asset-experience (periods 5, 7 and 8) yields average bids that appear slightly higher for dissimilar assets than for similar assets (580 for asset\_05=3 vs. 504 for asset\_05=2). This difference is not significant ( $p=0.28$ ). Similarly, the average bids in periods 14, 16 and 17 (the second new-asset experience), appear slightly higher for a

---

<sup>13</sup> That is, periodic assessment of gold reputation taken at the ends of periods 2,5,8,11,14 and 17 were highly correlated with probability estimates of assurer success taken at the beginning of the (gold) periods 3,6,9,12,15 and 18.

similar asset than for a dissimilar asset, but again they are not significantly so (377 for  $\text{asset\_05}=3$  vs. 340 for  $\text{asset\_05}=2$ ,  $p=0.51$ ). This pattern, though a statistically non-significant finding, is not inconsistent with the hypothesis that the assurance on a dissimilar asset *tends to* be a little more valuable in the short run (perhaps due to uncertainty with the asset), but in the long run the market values the higher accuracy (with the similar asset) more.

To investigate these possibilities better, a set of two Repeated Measures Analysis of Variance models was run, with either Bids 5 through 8 or Bids 14 through 17 as the dependent repeated measure, and  $\text{asset\_05}$  as the independent variable (capturing the difference between the top and bottom halves of the Figure 1 graph). A summary of the results is reported at the bottom of Figure 1. Interestingly, the models indicate a within subjects  $\text{asset\_05}$  interaction effect in the short run (i.e., in bids 5-8,  $\text{trials*asset\_05}$  interaction,  $p=0.004$ ) and a between subjects  $\text{asset\_05}$  effect in the long run (i.e., in bids 14-17,  $\text{asset\_05}$  effect,  $p=0.010$ ).<sup>15</sup> In the short run, when the first new asset is experienced, subjects change their bids differently (within themselves) based on which new asset is experienced first; however, the average bids are not significantly different between the two asset types. For the second new asset experienced, subjects do not change their bids (within subjects) in a differential manner based on which sequence they had experienced. However, in this case there is a significant difference in the average bids based on what sequence was experienced earlier. In other words, it appears that in the short run (first experience with a new asset), bids tend to remain static for a similar asset and fall for a dissimilar asset, but

---

<sup>14</sup> We present plots of medians for Figures 1, 2 and 3, but the plots of means are broadly consistent with these figures in each case.

<sup>15</sup> When subjects' probability estimates for these periods are added as covariates in the two models, the between-subjects effect is significant in both cases. Also see analysis of probability estimates for all 18 periods, later.

there is no difference in the average bid between these two cases; in the long run (second experience with a new asset), there is less of a slope difference, but the averages are different.

Analysis of Probability Estimates for All 18 Periods. Figure 2 presents the median probability estimates over all 18 periods, separated by the new asset sequence. Median probability estimates of assurer accuracy over periods 5, 7 and 8 (the first new-asset periods) are higher ( $p=0.000$ ) for a similar asset (mean 0.68 when  $\text{asset}_{05}=2$ ) as compared with a dissimilar asset (mean 0.57 when  $\text{asset}_{05}=3$ ). Correspondingly, over periods 14, 16 and 17 (the second new-asset periods), mean probability estimates still tend to be higher for similar assets than for dissimilar assets but not significantly so (0.58 for  $\text{asset}_{05}=3$  vs. 0.56 for  $\text{asset}_{05}=2$ ,  $p=0.33$ ). To investigate this further, two Repeated Measures Analysis of Variance models were run for probabilities as for bids. The results are summarized at the bottom of Figure 2.

It appears from the ANOVA models that there is a significant difference in how probability estimates change within subjects, based on whether the similar or dissimilar asset was experienced first ( $\text{trial}*\text{asset}_{05}$  interaction,  $p<0.01$  in both cases). As the graph suggests, for both the first experience (periods 5-8, upper half of graph) and the second experience (periods 14-17, lower half of graph) of a new asset, a similar asset causes a relatively flatter probability-estimate updating than a dissimilar asset. Conversely, a dissimilar asset causes a gradual decrease in probability estimates of assurer accuracy (especially when periods 6, and 15, the two gold periods, are omitted). But the between-subjects effect of  $\text{asset}_{05}$  (the sequence) on probabilities in periods 14 through 17 is non-significant ( $p=0.51$ ) unlike the corresponding effect for periods 5 through 8 ( $p=0.000$ ).<sup>16</sup> This is somewhat surprising on casual observation, since the opposite effect was observed for bids, in which

case the between subjects effect was significant for periods 14-17 but not for periods 5-8. However, this underscores the difference between bids and probability estimates as measures of reputation. While probability estimates measure participants' beliefs regarding the assurer's possible accuracy in the immediately following period, the bids measure a combination of the probability estimates and perceptions of value of the attestation service. More specifically, it already has been noted that probability estimates and bids do not correlate well in periods 4, 6, 7, 12 and 13. In periods 14, 15, 16 and 17, the correlations between probabilities and bids were 0.18, 0.13, 0.20 and 0.24 respectively ( $p < 0.01$  in all cases). Over the longer-term, participants' probabilistic beliefs about assurer accuracy qualitatively converged across the two new asset sequences, but different beliefs about attest report's value lingered across the two new asset sequences. As such, the inconsistency between probability estimates and bid measures was not unexpected.

Analysis of Periodic Assessments of Reputation for all 18 periods. Adopting a consumer-behavior orientation, reputation in our study was also measured by direct solicitation. Overall reputation, and reputation in the gold, silver and diamonds markets were collected separately at the end of periods 2, 5, 8, 11, 14 and 17 (see italicized in Exhibit 1, Panel C). Figure 3 presents the medians of these reputation measures. As would be expected, the assurer's reputation is always highest for the flagship service, followed by silver and diamond, in that order. Interestingly, subjects consistently measured overall reputation at a level consistently below the flagship reputation, showing the tainting effect of the new services concurrently experienced. Overall reputation is lower in the short run when the first new assurance experienced is dissimilar to the flagship service (periods 5-8,  $asset\_05=3$ ) as compared with when the new assurance experienced is similar to the flagship (periods 5-8,

---

<sup>16</sup> Omitting period 15 from the period 14-17 group reduces the p value from 0.51 to 0.33 (also see t-test

asset\_05=2, between subjects ANOVA  $p=0.000$ ). The data suggests that overall reputation drops within subjects in the short run with across a new dissimilar asset experience, while it remains fairly constant across a new similar asset (trials \* Asset\_05 interaction  $p=0.000$ ). This pattern is repeated for the second new asset (trials \* Asset\_05 interaction  $p=0.000$ ): overall reputation drops within subjects for the dissimilar asset (asset\_05=2, periods 14-17) but remains fairly constant for the similar asset (asset\_05=3, periods 14-17).

The elicited reputation for gold tells a reassuring story for flagship service reputation. Consistent with a hypothesis of encapsulation, there is no significant effect, between or within subjects ( $p>0.10$  for all tests in both short-term and long-term models). As the figure illustrates, reputation for gold remained relatively untainted by changes in levels of the independent variables both across the first experience (periods 5-8) and the second experience (periods 14-17) with a new service.

The reputation for silver also seems encapsulated. Reputation for the assurer's silver assurance ability goes up with the first experience in the silver service and remains largely constant (i.e., unaffected) across the first experience with the inferior-performance diamond service (periods 5-8, between subjects  $p=0.000$ , trials \* asset\_05 interaction  $p=0.032$ ). However, with the experience of a second new service (periods 14-17), the results are less straightforward. The reputation for silver is actually higher (between subjects  $p=0.000$ ) across concurrent experience with a dissimilar service with inferior performance (diamonds, periods 14-17, asset\_05=2), than with a concurrent experience of relatively superior performance in the silver service itself (periods 14-17, asset\_05=3). On the other hand, there is no significant trials \* Asset\_05 interaction ( $p=0.86$ ), suggesting that participants did *not* revise their silver reputation opinions *differentially* based on asset\_05 levels. These

---

reported earlier) but does not make the between-subjects effect statistically significant at conventional levels.

somewhat paradoxical findings for silver reputation in periods 14-17 can be explained by the fact that, going into period 14, silver reputation was already at a higher level for asset\_05 =2 as compared to asset\_05=3 due to the initial favorable (unfavorable) experience with silver (diamonds) in periods 5-8. Thus, the silver reputation data for periods 14-17 seem to indicate on the whole that while silver reputation does not get hurt by other service performance in the long run, it does not get enhanced by further success in silver either. In other words, after a reputation for silver (a new asset) has been established (e.g. with good performance in periods 5-8 and or transfer of reputation from the flagship), the reputation is difficult to improve upon, but also resilient to hurts caused by bad performance in *another* new service.

Reputation for the diamonds assurance service yielded data broadly consistent with the dynamics of the silver service reputation. In the short run (periods 5-8), diamonds service reputation was hurt by inferior performance in diamonds (asset\_05=3) but relatively unaffected across superior performance in silver (trials \* asset\_05 interaction,  $p=0.000$ ). In the long run (periods 14-17), inferior performance in the diamonds service (asset\_05=2) affected its reputation, and superior performance in silver (asset\_05=3) did not improve the diamond reputation (trials \* asset\_05 interaction,  $p=0.026$ ).

Detailed Analysis of Fully Crossed Part of Our Design. Since the periods starting with period 05 (i.e., when a first new asset is experienced) are extremely important in understanding the immediate (as opposed to long term) effects of introducing a new service, we analyzed these periods separately in further detail. Our experimental design lends itself to a close scrutiny, using fully crossed models, of the periods immediately preceding and following this first changeover from a flagship product to the other products. The following sections present these “micro” analyses.

Initial Transfer of Flagship Reputation to New Services. After three periods (00 to 02), participants had observed the assurer report correctly on the quality of gold (the flagship) three times. Given the differential overlap in knowhow, we expected that participants who rated the assurer's gold reputation higher would be more prone to also rate the assurer's silver reputation (in anticipation) higher than the assurer's diamond reputation. To test this expectation, we ran a repeated-measures ANOVA with the post-period 02 silver and diamond reputations as dependent measures and with the post-period 02 gold reputation as the independent measure. Our expectation would manifest as an interaction in which the difference between the silver and diamonds reputations increases with higher assessments of the gold reputation. This interaction is significant at ( $p_{\text{one-tailed}}=0.0593$ ). Subsequent univariate models show that the post-period 02 gold reputation is associated with the post-period 02 silver reputation at ( $p_{\text{one-tailed}}=0.0066$ ) and with the post-period 02 diamond reputation at ( $p_{\text{one-tailed}}=0.0590$ ). While these findings are consistent with a greater extent of transfer of flagship reputation to silver than to diamonds, it is noteworthy that the effect is not large (the R-squares in the univariate models were 0.021 and 0.008 for silver and diamonds, respectively).

Before trading in period 05, the first new-service period, market participants assessed the likelihood that the assurer would perform successfully and submitted their bids for the exclusive right to view the assurer's attest report. We expected whether the new service involved silver or diamonds would be more salient to participants who believed the silver and diamond reputations to differ markedly after period 02. These differences between silver and diamond reputation after period 02, as noted above, were increasing in participants' assessments of gold reputation after period 02. To examine for this relation, we estimated a general linear model using participants pre-period 05 probability assessments

(P\_05) as a dependent measure. For independent variables we included and crossed a categorical variable for new asset type (ASSET\_05 = silver or diamonds) and a continuous variable for the difference in participants' post-period 02 silver and diamond reputations (PA\_02SvD). Figure 4 plots the expected significant interaction between ASSET\_05 and PA\_02SvD ( $p_{\text{one-tailed}}=0.0022$ ) and provides the results of the general linear model. As shown, as participants perceived the silver reputation to exceed the diamond reputation post-period 02, they made higher (lower) probability assessments that the assurer would correctly attest to the quality of silver (diamonds).

We next estimated the same model using participants' pre-period 05 bids for the attest report (Bid\_05) instead of their assessed probabilities (P\_05) as the dependent measure. The anticipated interaction fails to attain significance ( $p_{\text{one-tailed}}=0.4305$ ). The same holds even when we add P\_05 to the model as an additional independent variable. While P\_05 explains a significant portion of the variation in Bid\_05 at ( $p_{\text{one-tailed}}=0.0001$ ), the ASSET\_05 by PA\_02SvD interaction remains insignificant ( $p_{\text{one-tailed}}<0.1258$ ).

To summarize the very short term results, evidence suggests that greater perceived knowhow overlap between flagship and new services facilitates the transfer of flagship reputation to the new service and that differential new service reputation affects market participants' assessments of the probability that the assurer will succeed in new attest services. The effects of differential new service reputation do not, however, have a significant effect on market participants' bids for the attest report. In hindsight, there are several possible explanations for the lack of complete mapping from assurer new-service reputations to bids for exclusive access to the assurer's attest report. One, probabilities of assurer success themselves explain only 4% of the variation in the bids for the assurer's attest report. In this sense it is not surprising that market participants' assessments of assurer's

new-service reputations, which explain around 5% of the variation in their assessed probabilities of assurer success, do not significantly affect bids. Two, because a number of market participants may have been attempting to win the assurer's attest report at all costs, other participants may have rationalized that any reasonable bid would not land them the attest report in any case. As such, they may not have exerted significant effort in determining the optimal bid amount, creating noise in bids. Because it is more difficult to calculate the expected value of winning the assurer's attest report than to estimate the probability of assurer success (the latter is an input into the former), it would be surprising if bids were not more noisy than probabilities.

Ramifications of New Service Performance for Flagship Service Reputation. Our next expectation is that performance in new service markets affects both overall reputation and, more interestingly, flagship service reputation. We expect that these effects would be stronger the greater the knowhow overlap between the flagship and new service. To investigate this relation, we first examined whether the post-period 05 overall (PA\_05O) and gold reputations (PA\_05GOL) were affected by success or failure by the assurer in period 05 and whether the new service involved silver (stronger knowhow overlap) or diamonds (weaker knowhow overlap). This analysis also controls for the actual quality of the asset traded in period 05, which could be either high or low quality. The only significant effect in the model is assurer success or failure at ( $p_{\text{one-tailed}}=0.0176$ ), as plotted and reported in Figure 5. Thus, market participants appear to have discounted the nature of the new service in assessing the flagship reputation implications of the assurer's new service performance.

The model concerning overall reputation presents similar findings, although the effect of assurer success or failure in the new service market is stronger ( $p_{\text{one-tailed}}<0.0001$ ), explaining about 25% of the variation in overall reputation as opposed to only about 3% of

the variation in flagship service reputation. Figure 6 presents a plot of the success versus failure effect and reports the general linear model.

We also examined whether the impact of success or failure in the new service market on the assurer's flagship reputation similarly affected the market participants' pre-period 06 probability assessments about whether the assurer would be correct as well as their bids for exclusive access to the assurer's attest report. To examine probabilities, we ran the same model except we substituted the P\_06 for the dependent variable for the periodic assessment of the assurer's reputation. Within this model, however, no effects are significant (lowest  $p_{\text{one-tailed}}=0.1400$ ), regardless of whether we include the last gold-related probability (P\_04) as a covariate. To examine bids, we substituted BID\_06 as the dependent measure. When prior bids (BID\_04) are not included as a covariate in the model, success versus failure is significant in the predicted direction at a generous alpha level ( $p_{\text{one-tailed}}<0.0688$ ).<sup>17</sup> However, when one includes prior bids, the significance of success goes away ( $p_{\text{one-tailed}}<0.1372$ ).

Overall, the very short term finding that the success or failure of the assurer in new service markets significantly affects the assurer's flagship reputation is inconsistent with encapsulation, while economic encapsulation is weakly suggested by other findings. Namely, neither market participants' probability assessments that the assurer would succeed or fail in the next period involving the flagship service nor their bids for exclusive access to the assurer's report is significantly affected by assurer success or failure in the new markets.

---

<sup>17</sup> The model of market participant's pre-period 06 bids also includes a significant ASSET\_05 effect that goes in the direction opposite to that which we predicted ( $p_{\text{two-tailed}}<0.0052$ ) as well as a significant main effect for quality ( $p_{\text{two-tailed}}<0.0432$ ). This ASSET\_05, however, appears to stem from a path dependency in that bids for markets in which diamonds were the first new service started off higher (for some reason) than did bids in markets in which silver was the first new service. The quality effect is such that higher bids prevail in period\_06 with higher quality assets traded in period 05. It is difficult to identify the reason behind the quality effect, but a salient post-hoc explanation is that it could stem from gambler's fallacy reasoning. As it turned out, when the quality was high in period 05, it was the third consecutive period of high quality. Participants may have reasoned that, if quality were low in period 06 and the (purchased) assurer report indicated so, their relative position in terms of francs, the experimental currency, could improve considerably by not purchasing assets while everyone else did so.

Consumer behavior research had found evidence of flagship encapsulation only when the fit between the product extension and the brand umbrella was low. Here we see some evidence of encapsulation despite a significant degree of fit between gold and silver attestation processes. Of course, the key question is what theoretical reason(s) explain why significant damage to the flagship reputation does not reliably affect probabilities of assurer success in the flagship service or bids for exclusive access to the assurer's attest report. A plausible explanation is that the initial significant flagship damage as a result of new service success and failure is itself an apparition that dissipates over repeated interactions. As market participants are provided with track records of the flagship performance and new service performance, they may come to learn that success or failure is not as predictive of performance in the flagship market, which, as operationalized, was perfect.

Persistence of Flagship Reputation Benefits and Damage Related to New Service Performance. To examine whether the assurer's longer-term track record of success or failure in new markets continues to influence the assurer's flagship reputation, we calculated several running success score variables. Different versions of this variable are applicable to the assurer's flagship reputation at different periods. Panel A of table 2 provides a summary of the statistical significance of success versus failure in the new service markets for the flagship reputation.<sup>18</sup> As shown, the damage of the flagship reputation based on success or failure in new assurance markets is a persistent effect, but also one that explains a small

---

<sup>18</sup> A repeated measures ANOVA including all six measures of the flagship gold reputation yields similar results with a ( $p_{\text{one-tailed}} < 0.0445$ ). More complex models including ASSET\_05 and a running tally of quality yield no further significant results. It is noteworthy that the design is not fully crossed after period 5 in the market sessions due to stochastic realizations of asset quality and assurer success or failure, limiting the types of tests we can perform to identify the asset by success interaction we hypothesize based on the consumer behavior literature.

amount of the variation in the flagship reputation (averaging around 1% of explained variation).<sup>19</sup>

Panel B provides the same set of analyses for the assurer's overall reputation as a point of comparison. In contrast to the assurer's flagship reputation, the assurer's overall reputation is strongly affected by the degree of success realized in new assurance services. Across the six periodic assessments of assurer reputation, the assurer's cumulative success or failure in new services to date explains an average of 19.3% of the variation in overall reputation.

These findings suggest that, although the effect size is small, the effects of new service success and failure persistently affect the assurer's flagship reputation. This finding is consistent with consumer behavior research on brand extensions holding that consumer attitudinal evaluations of product extensions affect the brand umbrella overall and, as a consequence, all products under that umbrella (Aaker and Keller [1990]). The underlying psychological mechanism for this effect likely is an associative memory and spreading activation (Collins and Loftus [1975]). Because the systematic effects on flagship reputation do not map into systematic economically relevant behaviors, the involuntary, memory-based explanation has appeal. Because economic ramifications of the flagship damage are not systematically observed in our experimental setting, the possibility arises that damage to a flagship reputation can be accompanied by flagship service encapsulation for economic purposes.

Of course, the key to verifying or refuting such a conjecture is to identify theoretically appealing reasons as to why attitudinal flagship reputations can suffer damage without systematically affecting measures such as the probability of assurer success and the

---

<sup>19</sup> Flagship probability assessments and bids were not significantly affected by assurer success or failure in the

amount one bids for exclusive access to the assurer's report. We tried to design our experiment so that participants were well aware of the assessments of assurer reputation. This approach was desirable so as to give reputation a good shot at explaining variation in probabilities and bids. But, of course, we cannot rule out the possibility that we fail to observe a systematic effect in probabilities and bids simply because of noise in these assessments.<sup>20</sup>

We can advance a post-hoc explanation for significant noise in bids. Early on it might have been the case that market participants were trying to learn the value of securing the assurer's attest report.<sup>21</sup> It also could have been the case that the "smart" participants realized that attest report winners paid far too much for it.<sup>22</sup> If this were true, these smart participants might have decided that intense deliberation over the amount to bid was of little value to them, creating noise in bids. In our experimental setting, market participants were not directly rewarded for making rational bids on assurer reports. Instead, they were rewarded in francs (convertible into lottery tickets) for overall trading profits less any francs they paid for access to assurer reports. It remains possible, therefore, that providing a more direct incentive to make optimal bids could enable future researchers to observe a systematic link between flagship reputation damage and bids for a flagship service.

## *5. Conclusion*

In our examination of the dynamics and economic implications of an assurance provider's (assurer's) reputation within a multi-service laboratory market setting, several interesting findings emerge. Consistent with our first hypothesis, we observe both

---

new markets.

<sup>20</sup> As of this draft, we have not completed sensitivity analyses involving standardizing participants bids in diverse ways or eliminating outlier or illogical bids or probabilities.

<sup>21</sup> Pre-market written and oral instructions explained and illustrated the concept of expected value for participants.

feedforward and feedback reputation effects. Flagship reputations feed forward to market participants' assessments of the assurer's new service reputation, and success and failure in new markets feeds back to affect the flagship reputation. Unexpectedly, the effects of success and failure do not reliably differ across the two new assurance service markets. It seems that a bounded rationality or a cost-of-processing argument could explain this finding, in hindsight. In order for the nature of the new assurance service to matter, market participants would have to engage in configural information processing, which requires effort. The benefits of exerting such effort may not have been apparent or may have been small within each market period. That is, bounded rationality might prohibit integration of the incremental diagnosticity of knowhow overlap above that of mere success or failure in new markets.

Further, extending the consumer behavior literature, we find that the feedback loops persist over repeated interactions. A plausible reason that these effects subsist is that they are so small that they do not affect bids, making them of little economic consequence. That is, despite the persistence of feedback loops, we find mixed evidence that spillover-related changes in assurer reputation influence likelihood assessments about assurer performance. While we find that market participants' initial probability assessments that the assurer would be successful in the new service is related to initial flagship reputation transfer, we do not observe feedback effects from new services to flagship services at the level of probabilities. Finally, we do not observe a systematic link between assurer performance in new services and participants' bids for the assurer's attest report.

Although more theory development is needed, the empirical observation of no relation between spillover effects that affect flagship reputation and actual probabilities of

---

<sup>22</sup> Although the winner's curse obtains, winners of the attest report earned more, on average, after adding back

assurer success or bids for flagship services is interesting. Conventional wisdom holds that reputation is hard to earn but easily lost (Slovic [1993]). Yaniv and Kleinberger (2000) put it like this: "Reputation building may be likened to biking uphill; that is, it is hard to start forward (gain) but easy to roll backward (lose)." Our study, in contrast, suggests that, at least in the absence of assurer competition, poor performance in related, new services does not substantially damage the economic rents earned from a flagship reputation.

## References

- Aaker, D. A., and K. L. Keller. 1990. Consumer evaluations of brand extensions. *Journal of Marketing* 54 (January): 27-41.
- Ackert, L.F., Church, B. K. and Zhang, P. 2001. Forecast characteristics, market prices, and prediction biases: evidence from experimental asset markets. Working Paper.
- AICPA. 1999. Report of the Special Committee on Assurance Services. <http://www.aicpa.org/assurance/scas/index.htm>. (12 July 2000).
- Bandyopadhyay, S. P., and J. L. Kao. 1999. Competition and Big 6 brand name reputation: Evidence from the Ontario municipal audit market. Working paper. University of Waterloo and University of Alberta.
- Banerjee, A. V., and E. Duflo. 2000. Reputation effects and the limits of contracting: A study of the Indian software industry. *The Quarterly Journal of Economics* 115 (August): 989-1017.
- Bloomfield, R. and J. Hales. 2001. Predicting the next step of a random walk: Experimental evidence of regime-shifting beliefs. Working paper, Cornell University.
- Boritz, J. E. and D. Cockburn. 1998. 1998 Audit Symposium panel discussion on assurance services. *Auditing: A Journal of Practice and Theory* 17 (Supplement): 133-151.
- Broniarczyk, S. M. and J. W. Alba. 1994. The importance of the brand in brand extension. *Journal of Marketing Research* 31 (May): 214-228.
- Camerer, C. 1981. *The validity and utility of expert judgment*. Ph.D. dissertation, University of Chicago.
- Camerer, C. and K. Weigelt. 1988. Experimental tests of a sequential equilibrium reputation model. *Econometrica* 56 (January): 1-36.
- Carpenter, P. A., M. A. Just, and P. Shell. 1990. What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review* 97 (no. 3): 404-431.
- Cattell, R. B. 1963. Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology* 54: 1-22.
- Charness, G. and N. Garoupa. 2000. Reputation, honesty, and efficiency with insider information: An experiment. *Journal of Economics and Management Strategy* 9 (Fall): 425-451.
- Chu, W. and W. Chu. 1994. Signaling quality by selling through a reputable retailer: An example of renting the reputation of another agent. *Marketing Science* 13 (Spring): 177-189.

- Clark, B. H. and D. B. Montgomery. 1998. Deterrence, reputations, and competitive cognition. *Management Science* 44 (January): 62-82.
- Collins, A. and E. F. Loftus. 1975. A spreading activation theory of semantic processing. *Psychological Review* 82: 407-428.
- Dacin, P. A. and D. C. Smith. 1994. The effect of brand portfolio characteristics on consumer evaluations of brand extensions. *Journal of Marketing Research* 31 (May): 229-242.
- Datar, S. and M. Alles. 1999. The formation and role of reputation and litigation in the auditor-manager relationship. *Journal of Accounting, Auditing, and Finance* 4: 401-428.
- Dhar, R. and R. Glazer. 1996. Similarity in context: Cognitive representation and violation of preference and perceptual invariance in consumer choice. *Organizational Behavior and Human Decision Processes* 67 (September): 280-293.
- Dopuch, N. and R. R. King. 1991. *Experimental tests of auditing as a credibility generating mechanism*. Research Monograph No. 19. The Canadian General Accountants' Research Foundation.
- Einhorn, H. J. 1980. Learning from experience and suboptimal rules in decision making. In *Cognitive Processes in Choice and Decision Behavior*, T. Wallsten (Ed.), Hillsdale, NJ: Lawrence Erlbaum.
- Elliot, R. K. 1997. Assurance service opportunities: Implications for academia. *Accounting Horizons* 11 (December): 61-74.
- Elliot, R. K. 1998. Assurance services and the audit heritage. *Auditing: A Journal of Practice and Theory* 17 (Supplement): 1-7.
- Evans, R. and J. P. Thomas. 1997. Reputation and experimentation in repeated games with two long-run players. *Econometrica* 65 (September): 1153-1173.
- Ferguson, T. D., D. L. Deephouse, and W. L. Ferguson. 2000. Do strategic groups differ in reputation? *Strategic Management Journal* 21: 1195-1214.
- Firth, M. 1990. Auditor reputation: The impact of critical reports issued by government inspectors. *RAND Journal of Economics* 21 (Autumn): 374-387.
- Fredeen, B. 2000. Fool's den: Know your brand. <http://www.fool.com/research/2000/foolsden000829.htm>. (May 6, 2001).
- Fudenberg, D. and D. M. Kreps. 1987. Reputation in the simultaneous play of multiple opponents. *Review of Economic Studies* 54: 541-568.

- Ganguly, A.R., J.H. Kagel, and D.V.Moser. 1994. The effects of biases in probability judgments on market prices. *Accounting, Organizations and Society*, 19 (8): 675-700
- Ganguly, A.R., J.H. Kagel, and D.V.Moser. 2000. Do Asset market Prices Reflect Traders' Judgment Biases? *Journal of Risk and Uncertainty*, 20(3): 219-245
- Hertwig, R. and A. Ortman. 2001. Experimental practices in economics: a methodological challenge for psychologists? *Behavioral and Brain Sciences*, 24(4), forthcoming.
- Jackson, M. O. and E. Kalai. 1999. Reputation versus social learning. *Journal of Economic Theory* 88: 40-59.
- John, D. R., B. Loken, and C. Joiner. 1998. The negative impact of extensions: Can flagship products be diluted? *Journal of Marketing* 62 (January): 19-32.
- Jung, Y. J., J. H. Kagel, and D. Levin. 1994. On the existence of predatory pricing: An experimental study of reputation and entry deterrence in the chain-store game. *RAND Journal of Economics* 25 (Spring): 72-93.
- Kagel, J. H. 1995. Auctions: A survey of experimental research. In *The Handbook of Experimental Economics*. J. H. Kagel and A. E. Roth (Eds): Princeton: Princeton University Press.
- Keller, K. L. 1993. Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing*. 57 (January): 1-22.
- King, R. R. 1996. Reputation formation for reliable reporting: An experimental investigation. *The Accounting Review* 71 (July): 375-396.
- King, R. R., and R. Schwartz. 1998. Planning assurance services. *Auditing: A Journal of Practice & Theory*. 17 (Supplement): 9-36.
- Klein, D. B. 1999. Trust for hire: Voluntary remedies for quality and safety. In *Reputation: Studies in the Voluntary Elicitation of Good Conduct*, edited by D. B. Klein. Ann Arbor: The University of Michigan Press, 97-133.
- Loken, B., and D. R. John. 1993. Diluting brand beliefs: When do brand extensions have a negative impact? *Journal of Marketing* 57 (July): 71-84.
- Mayhew, B. W. 2001. Auditor reputation building. University of Wisconsin-Madison Working Paper.
- Medin, D. L., R. L. Goldstone, and D. Gentner. 1993. Respects for similarity. *Psychological Review* 100 (April): 254-278.
- Neral, J. and J. Ochs. 1992. The sequential equilibrium theory of reputation building: A further test. *Econometrica* 60 (September): 1151-1169.

- Park, E. 1999. Reputation effects in bargaining games. *Southern Economic Journal* 66: 186-192.
- Plott, C. R. and P. Gray. 1990. Multiple Unit Double Auction. *Journal of Economic Behavior and Organization* 13: 245-258.
- Raven, J., J. C. Raven, and J. H. Court. 1995. *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1: General Overview*. Oxford: Oxford Psychologists Press, 1995.
- Shocker, A. D., R. K. Srivastava, and R. W. Rueker. 1994. Challenges and opportunities facing brand management: An introduction to the special issue. *Journal of Marketing Research* 31 (May): 149-158.
- Slovic, P. 1993. Perceived risk, trust, and democracy: A systems perspective. *Risk Analysis* 13: 675-682.
- Simonson, I. and A. Tversky. 1992. Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research* 29 (August): 281-295.
- Sullivan, M. 1990. Measuring image spillovers in umbrella-branded products. *Journal of Business* 63 (no. 3): 309-329.
- Sundem, G., R. Dukes and J. Elliott. 1996. *The Value of Information and Audits*. Coopers & Lybrand, 1996.
- Teece, D. 1980. Economies of scope and the scope of the enterprise. *Journal of Economic Behavior and Organization* 1 (September): 223-247.
- Teece, D. 1982. Towards an economic theory of the multiproduct firm. *Journal of Economic Behavior and Organization* 3 (March): 39-63.
- Underwriters Laboratory, Inc. 2000. [www.ul.com](http://www.ul.com) (12 July 2000).
- Vickrey, W. 1961. Counterspeculation, auctions and competitive sealed tenders. *Journal of Finance* 16: 8-37.
- Watts, R. and J. Zimmerman. 1983. Agency problems, auditing and the theory of the firm: Some evidence. *Journal of Law and Economics* 26 (October): 613-634.
- Wilson, T. E., and R. A. Grimlund. 1990. An examination of the importance of an auditor's reputation. *Auditing: A Journal of Practice and Theory* 9 (Spring): 43-59.
- Wyer, R. S., Jr. and T. K. Srull. Person memory and judgment. *Psychological Review* 96 (1): 58-83.
- Yaniv, I. and E. Kleinberger. 2000. Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processing* 83 (November): 260-281.



## Exhibit 1

### Details of the Experimental Market Sessions

#### Panel A: Timeline of experiment

Prior to experimental markets session:

1. Practice second-price, sealed-bid auction
2. Three-section quiz
3. Follow-up/manipulation check questions regarding quiz

During experimental markets session:

4. Instructions
5. Practice session
6. Experimental markets session
  - For each market period:
    - a. Experimenters announce asset type being sold in that period (gold, silver, or diamonds)
    - b. Conduct second-price, sealed-bid auction for assurer's report
    - c. Conduct double auction for asset
    - d. Experimenters announce asset quality
    - e. Subjects calculate profits for period, place on profit report
    - f. Periodically (every third period), subjects fill out a periodic assessment of assurer's reputation
7. Calculate total profits on profit report
8. Post market questionnaire/manipulation checks

#### Panel B: Relationship between quiz and assurer's report

The accuracy of the assurer's report on:

Depends on the top students' performance on:

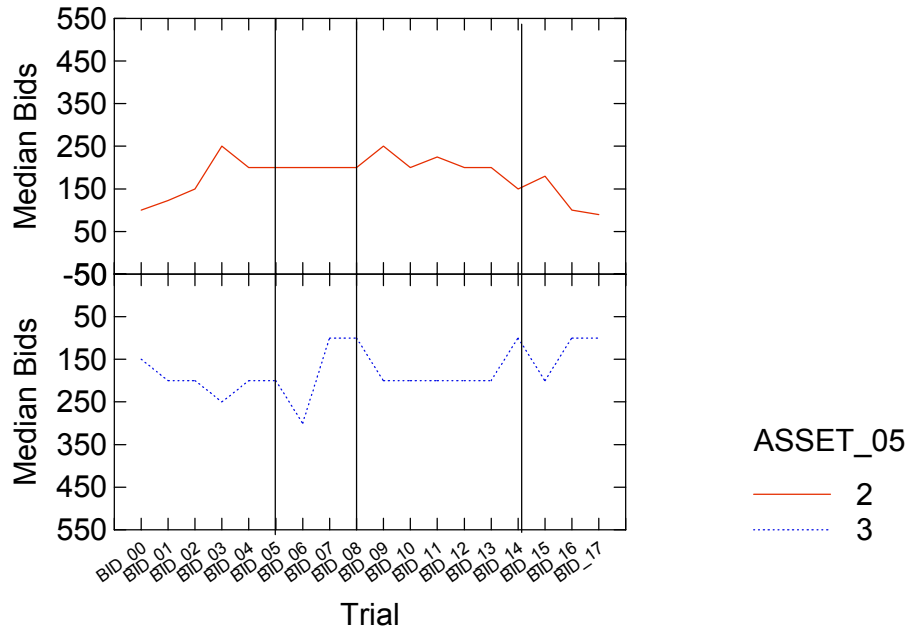
Gold	Analytical section
Silver	Financial-Analytical section
Diamonds	Financial-Recall section

#### Panel C: Order of assets traded

Period	Asset traded
00	Gold
01	Gold
<i>02</i>	<i>Gold</i>
03	Gold
04	Gold
<i>05</i>	<i>Silver/Diamonds</i>
06	Gold
07	Silver/Diamonds
<i>08</i>	<i>Silver/Diamonds</i>
09	Gold
10	Gold
<i>11</i>	<i>Gold</i>
12	Gold
13	Gold
<i>14</i>	<i>Diamonds/Silver</i>
15	Gold
16	Diamonds/Silver
<i>17</i>	<i>Diamonds/Silver</i>
18	Gold

Note: Before each period, participants provided probabilities of an accurate assurer attest report and second-price, sealed-bids for exclusive access to the attest report. After every third period, they provided a set of periodic assessments of the assurer's overall, gold, silver, and diamond reputation (as indicated by the *italicized fonts* above).

**Figure 1**  
**Median Bids for Attest Report, Separated by New Asset Sequence**



**Results of Repeated Measures Analysis of Variance**

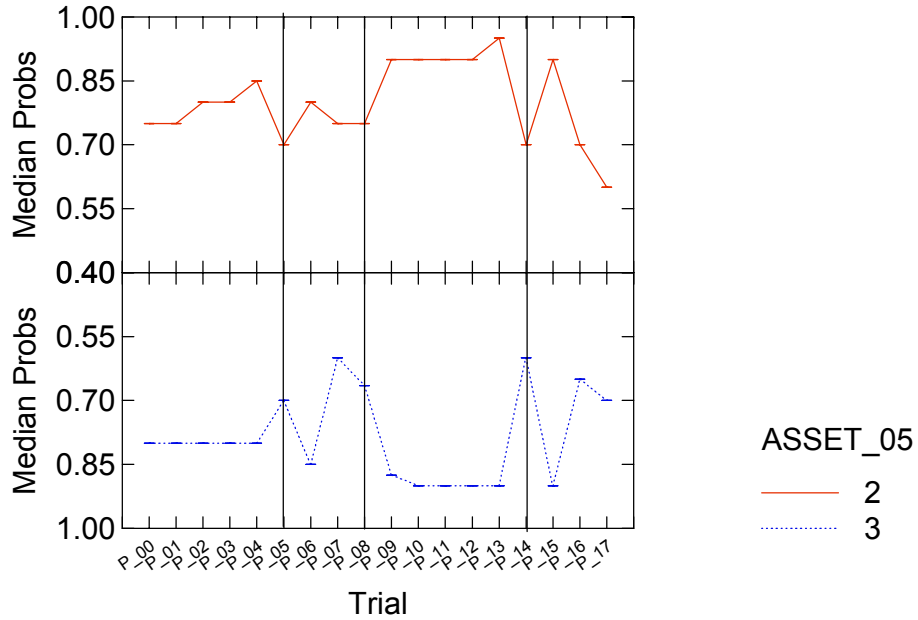
Summarized results of two models

Dependent Var	Between-Subjects		Within-Subjects	
	Source	p	Source	p
Bids 5-8	asset_05	0.132	trials	0.001
			trials*asset_05	0.004
Bids 14-17	asset_05	0.010	trials	0.000
			trials*asset_05	0.867

Asset\_05 refers to whether the order of new assurance markets is (2) higher knowhow overlap first, lower knowhow overlap second (i.e., silver first, diamonds second) or (3) lower knowhow overlap first, higher knowhow overlap second (i.e., diamonds first, silver second).

Also, on the lower half of the graph above, the vertical axis has been reversed so as to show the two sets of data as mirror images. Thus, on the lower half of the graph, increases in bids for Asset\_05=3 (diamonds first, silver second) are shown as downward sloping lines.

**Figure 2**  
**Median Probability Estimates of Assurer Accuracy, Separated by New Asset Sequence**



**Results of Repeated Measures Analysis of Variance**

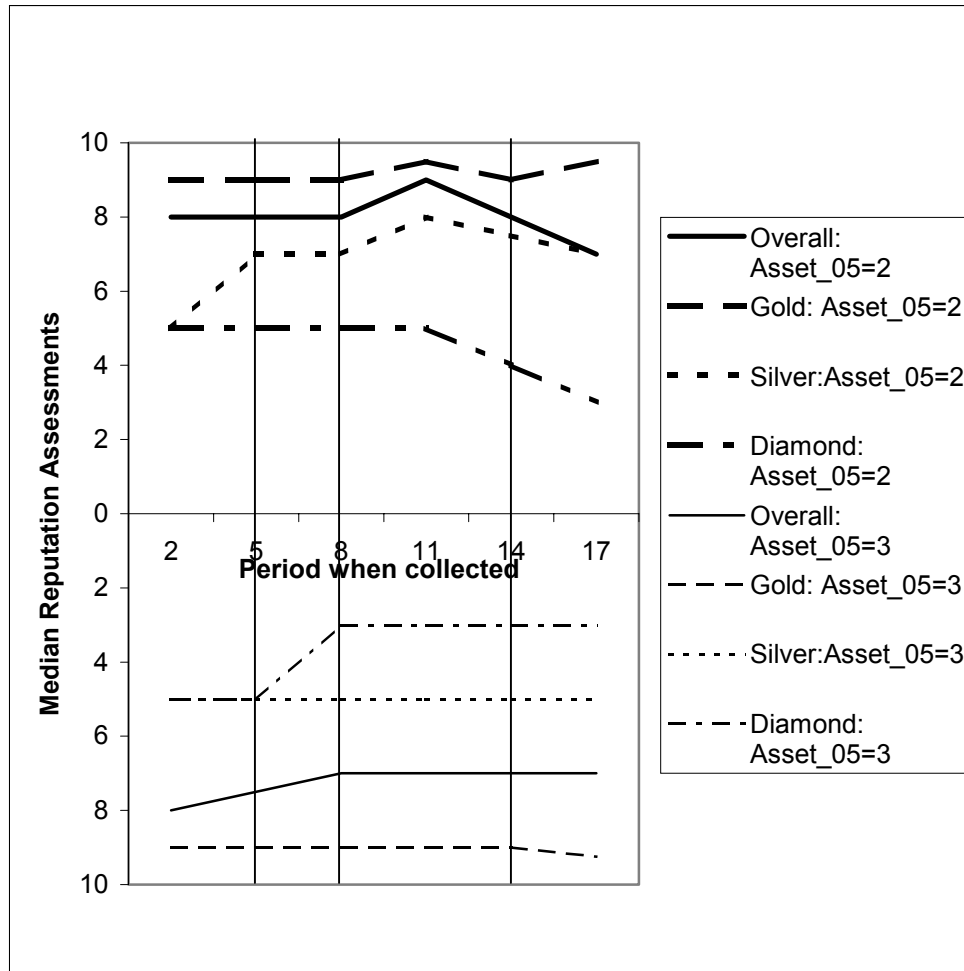
Summarized results of two models

Dependent Var	Between-Subjects		Within-Subjects	
	Source	p	Source	p
Probs 5-8	asset_05	0.000	trials	0.000
			trials*asset_05	0.000
Probs 14-17	asset_05	0.506	trials	0.000
			trials*asset_05	0.003

Asset\_05 refers to whether the order of new assurance markets is (2) higher knowhow overlap first, lower knowhow overlap second (i.e., silver first, diamonds second) or (3) lower knowhow overlap first, higher knowhow overlap second (i.e., diamonds first, silver second).

Also, on the lower half of the graph above, the vertical axis has been reversed so as to show the two sets of data as mirror images. Thus, on the lower half of the graph, increases in bids for Asset\_05=3 (diamonds first, silver second) are shown as downward sloping lines.

**Figure 3**  
**Median Periodic Reputation Assessments, Separated by New Asset Sequence (Asset\_05)**



**Results of Repeated Measures Analysis of Variance**

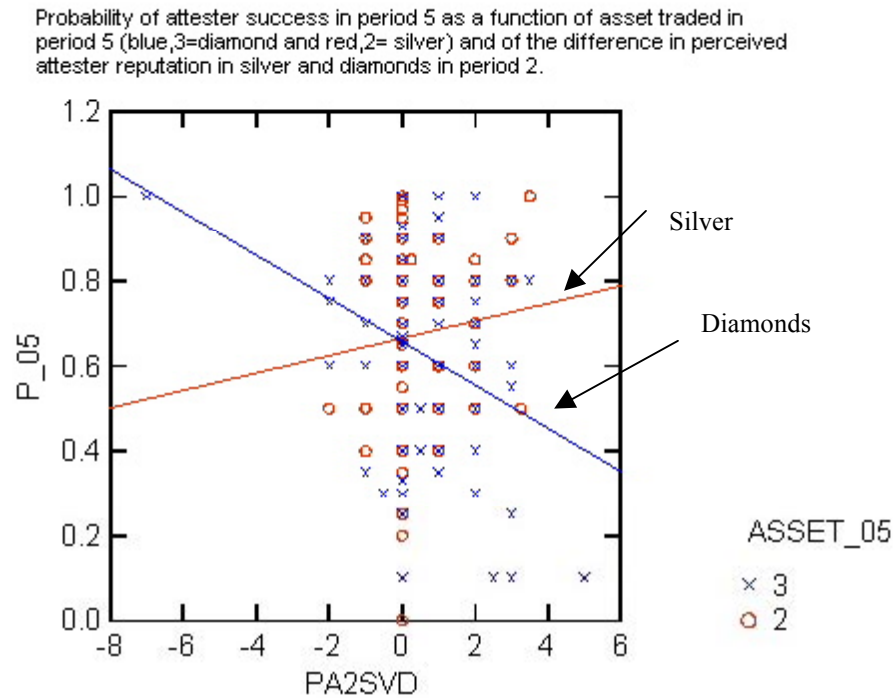
Summarized results of eight models

Dependent Variable	Between Subjects		Within Subjects			
	Source	p	Source	p	Source	p
Overall PA5-PA8	Asset_05	0.000	Trials	0.002	Trials * asset_05	0.000
Overall PA14-PA17	Asset_05	0.036	Trials	0.015	Trials * asset_05	0.000
Gold PA5-PA8	Asset_05	0.284	Trials	0.458	Trials * asset_05	0.316
Gold PA14-PA17	Asset_05	0.661	Trials	0.475	Trials * asset_05	0.309
Silver PA5-PA8	Asset_05	0.000	Trials	0.793	Trials * asset_05	0.032
Silver PA14-PA17	Asset_05	0.000	Trials	0.072	Trials * asset_05	0.860
Diamonds PA5-PA8	Asset_05	0.000	Trials	0.000	Trials * asset_05	0.000
Diamonds PA14-PA17	Asset_05	0.472	Trials	0.001	Trials * asset_05	0.026

Asset\_05 refers to whether the order of new assurance markets is (2) higher knowhow overlap first, lower knowhow overlap second (i.e., silver first, diamonds second) or (3) lower knowhow overlap first, higher knowhow overlap second (i.e., diamonds first, silver second).

Also, on the lower half of the graph above, the vertical axis has been reversed so as to show the two sets of data as mirror images. Thus, on the lower half of the graph, increases in bids for Asset\_05=3 (diamonds first, silver second) are shown as downward sloping lines.

**Figure 4**  
**Pre-period 05 Probability of Assurer Success as a Function of Asset Subjected to Attestation in the New Service (Asset\_05) and the Difference in Participants' Post-period 02 Assessments of Assurer Reputations for Silver and Diamond (PA2SVD)**



Note: In the plot above probabilities of assurer success for diamond (silver) attest reports are represented by x's and o's, respectively. The positively (negatively) sloped lines in the figure represent univariate regression lines for markets in which silver (diamonds) attestation was the first new assurance service.

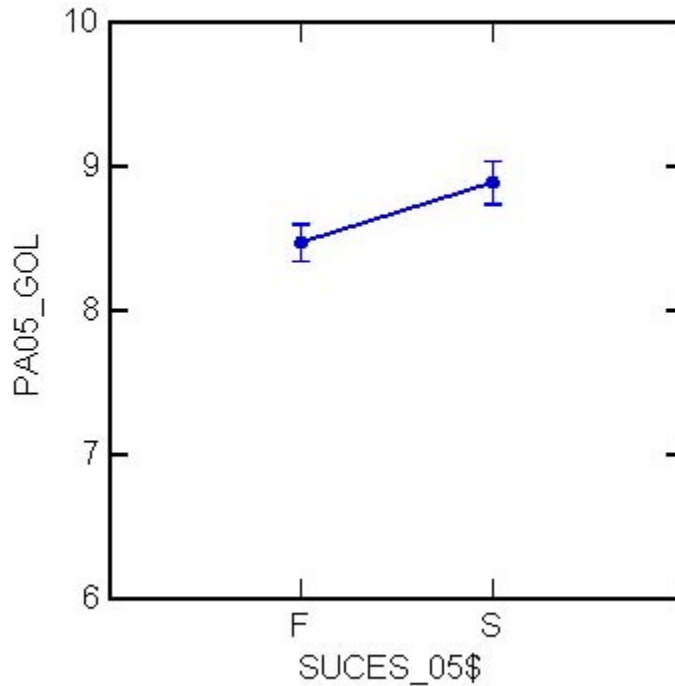
Dep Var: P\_05    N: 300    Multiple R: 0.2201    Squared multiple R: 0.0484

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
ASSET_05	0.0039	1	0.0039	0.0915	0.7625
PA2SVD	0.0631	1	0.0631	1.4913	0.2230
<b>ASSET_05*PA2SVD</b>	<b>0.3483</b>	<b>1</b>	<b>0.3483</b>	<b>8.2318</b>	<b>0.0044</b>
Error	12.5243	296	0.0423		

**Figure 5**  
**Post-Period 05 Flagship Service Reputation (PA\_05GOL) as a Function of the Asset Subjected to Attestation in the New Service (ASSET\_05), the Success or Failure in Assurer Performance (SUCES\_05), and the Quality of Assets Traded (QUAL\_05)**

**Least Squares Means**



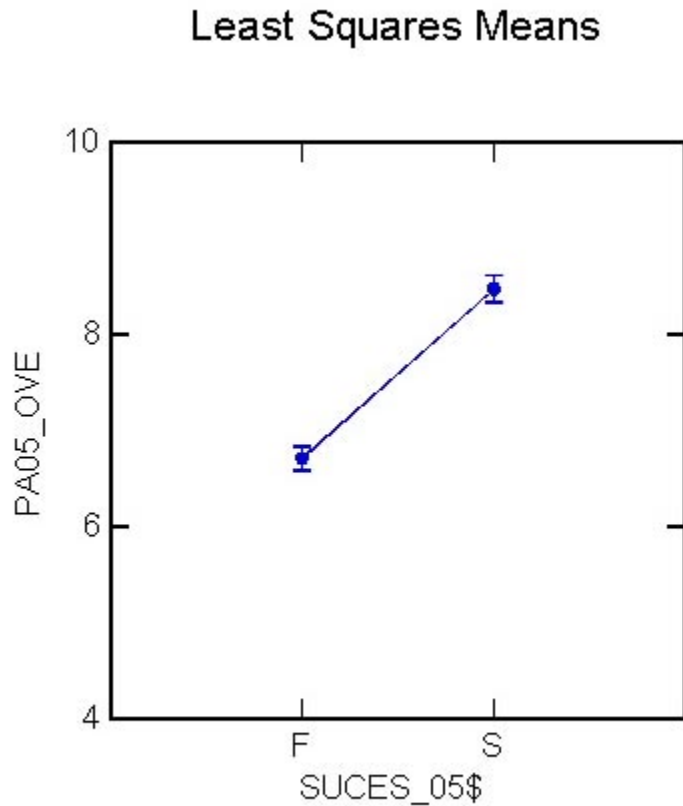
Note: PA05\_GOL is the post period 05 assessment of the assurer's gold attestation reputation, and SUCES\_05\$ is a category failure indicating failure (F) or success (S) in the new service market. The significant main effect success and insignificant asset by success interaction indicates that post-period 05 assurer flagship reputation increases with success and decreases with failure in new service, without prejudice to the extent of knowhow overlap between a flagship and new service.

Dep Var: PA05\_GOL N: 319 Multiple R: 0.1780 Squared multiple R: 0.0317

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
<b>SUCES_05</b>	<b>11.4228</b>	<b>1</b>	<b>11.4228</b>	<b>4.4816</b>	<b>0.0351</b>
QUAL_05	1.4977	1	1.4977	0.5876	0.4439
ASSET_05	0.3725	1	0.3725	0.1461	0.7025
SUCES_05*QUAL_05	1.4661	1	1.4661	0.5752	0.4488
SUCES_05*ASSET_05	4.1635	1	4.1635	1.6335	0.2022
QUAL_05*ASSET_05	0.0622	1	0.0622	0.0244	0.8760
SUCES_05*QUAL_05*ASSET_05	6.8195	1	6.8195	2.6756	0.1029
Error	792.6818	311	2.5488		

**Figure 6**  
**Post-Period 05 Overall Reputation (PA\_05OVE) as a Function of the Asset Subjected to Attestation in the New Service (ASSET\_05), the Success or Failure in Assurer Performance (SUCES\_05), and the Quality of Assets Traded (QUAL\_05)**



Note: PA05\_OVE is the post period 05 assessment of the assurer's overall reputation, and SUCES\_05\$ is a category failure indicating failure (F) or success (S) in the new service market. The significant main effect success and insignificant asset by success interaction indicates that post-period 05 assurer flagship reputation increases with success and decreases with failure in new service, without prejudice to the extent of knowhow overlap between a flagship and new service.

Dep Var: PA05\_OVE N: 319 Multiple R: 0.4951 Squared multiple R: 0.2451

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
ASSET_05	1.7675	1	1.7675	0.7527	0.3863
SUCES_05	203.3024	1	203.3024	86.5820	2.24160E-11
QUAL_05	4.0925	1	4.0925	1.7429	0.1877
ASSET_05*SUCES_05	2.1404	1	2.1404	0.9115	0.3404
ASSET_05*QUAL_05	0.7261	1	0.7261	0.3092	0.5786
SUCES_05*QUAL_05	0.1184	1	0.1184	0.0504	0.8225
ASSET_05*SUCES_05*QUAL_05	0.0314	1	0.0314	0.0134	0.9081
Error	730.2562	311	2.3481		

**Table 1**  
**Variance Explained in Overall Reputation**  
**Explained by Gold, Silver, and Diamond Reputations**

	PA_02	PA_05	PA_08	PA_11	PA_14	PA-17
Gold, Silver, & Diamonds	<b>0.50</b>	<b>0.61</b>	<b>0.60</b>	<b>0.43</b>	<b>0.51</b>	<b>0.50</b>
Silver & Diamonds	<b>0.05</b>	<b>0.31</b>	<b>0.44</b>	<b>0.28</b>	<b>0.33</b>	<b>0.25</b>
Gold	<b>0.46</b>	<b>0.37</b>	<b>0.30</b>	<b>0.21</b>	<b>0.24</b>	<b>0.34</b>
Silver	<b>0.05</b>	<b>0.29</b>	<b>0.43</b>	<b>0.28</b>	<b>0.32</b>	<b>0.24</b>
Diamonds	<b>0.05</b>	<b>0.16</b>	<b>0.17</b>	<b>0.16</b>	<b>0.07</b>	<b>0.05</b>

Adjusted R-squareds are shown

Significance Level of Models

p > .10

**p < .05**

**p < .001**

The periodic assessments (PA's) of assurer reputation first were made after the third period (i.e., period\_02 since the first period was period\_00) and then after every third period thereafter. In all six sets of periodic assessments exist, gathered after periods 02, 05, 08, 11, 14, and 17 (see Exhibit 1 for more details).

**Table 2**  
**Relation between Assurer Cumulative Success vs. Failure in New Service Markets**  
**and the Assurer's Flagship Reputation at Different Market Period**

## Panel A

	PA_05Gold	PA_08Gold	PA_11Gold	PA_14Gold	PA17_Gold	PostTestGold
p(one-tail)	0.0176	0.0177	0.0872	0.0247	0.0834	0.0834
success var	Suces_05	Suces_early	Suces_early	Suces_med	Suces_tot	Suces_tot
R-squared	1.4%	1.4%	0.6%	1.3%	0.8%	0.8%

## Panel B

	PA5OVE	PA8 OVE	PA11 OVE	PA14 OVE	PA17 OVE	PostTestOVE
p(one-tail)	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
success var	Suces_05	Suces_early	Suces_early	Suces_med	Suces_tot	Suces_tot
R-squared	23.3%	25.7%	21.0%	19.9%	10.1%	15.7%

Note: PA\_XXGold is the market participant's subjective assessment of the assurer's reputation for gold attestation services. Each periodic assessment of assurer reputation occurred after the trial indicated (i.e., PA\_05Gold was collected post-trading in period 05).

PAXOVE is the market participant's subjective assessment of the assurer's overall reputation collected after period X, where X=5, 8, 11, 14, 17.

PostTestGold and PostTestOVE is the market participant's subjective assessment of the assurer's reputation for gold and overall reputation, respectively, according to the post-market questionnaire, which was completed after they had double-checked their tabulation trading gains and losses in francs.

Suces\_05 is whether the assurer was correct in the first new service period.

Suces\_early is the summation of assurer success in new service periods 5, 7, and 8.

Suces\_med is the summation of assurer success in new service periods 5, 7, 8, and 14.

Suces\_tot is the summation of assurer success in new service periods 5, 7, 8, 14, and 17.