RAND Journal of Economics
Vol. 36, No. 2, Summer 2005
pp. 391–411

Cheap-talk referrals of differentiated experts in repeated relationships

In-Uck Park*

The effectiveness of cheap talk advice is examined in recurrent relationships between a customer and multiple experts who provide differentiated professional services. The main findings are: (i) Full honesty is not sustainable if the profitability of service provision varies widely across problems. (ii) As there are more experts due to finer specialization, the maximum equilibrium honesty level deteriorates. (iii) Nonetheless, the equilibria that satisfy an internal consistency condition, implement the same (unique) honesty level regardless of the number of experts. Furthermore, the customer can extract this honesty level by consulting a "panel" of only one or two (but no more) experts all the time.

1. Introduction

■ When patients see a doctor, they are often referred to another doctor who is better suited to treat the problem at hand. To reduce inefficiencies arising from patients' own choice of doctors, some health service providers require patients to consult initially with a designated doctor, known as the "gatekeeper," who will refer them to other doctors as necessary. Such schemes raise a number of incentive issues.[1] Similarly, large companies tend to seek advice on who can best handle various legal problems from in-house attorneys or law firms that they have on retainer. These counselors may either handle specific legal issues themselves or refer them to other more appropriate experts. Diagnosis by experts who may either treat the problem themselves or refer the customer to other service providers also arises in various types of consulting/advisory services and in repair services of durable goods such as houses and automobiles.

As these examples illustrate, a prominent feature of professions with differentiated specialties is that the customer often has to rely on experts' advice to identify the right service provider for each problem she faces. While matching the needed services to the right specialists increases the customer's surplus (and hence social welfare), the experts themselves have no intrinsic interest in giving truthful advice if their advice has no effect on future opportunities for providing service.

* University of Bristol, U.K.; I.Park@bristol.ac.uk.

I am grateful to the Editor Raymond Deneckere, Andrew McLennan, Jack Ochs, Tuvana Pastine, and a referee for helpful comments, and also to participants at seminars at Michigan State University, Rutgers University, and the University of Pittsburgh as well as at the New York Meeting of the Econometric Society (1999), the Southeast Economic Theory Conference in Washington, D.C. (1999), and the ESRC Research Seminars in Game Theory in Kennilworth, U.K. (1999). The research support of the Economic and Social Research Council (ESRC), U.K., is gratefully acknowledged (grant no. R000222716).

[1] Health maintenance organizations (HMOs) and Britain's National Health Service, among others, use gatekeeper arrangements. See, for example, Croxson, Propper, and Perkins (2001) for empirical findings on incentive issues of gatekeepers, and Malcomson (2004) for contract theoretic analysis.

Therefore, the customer who seeks advice from an expert who is also a service provider must be able credibly to threaten that expert with the loss of future business if the advice turns out to be wrong. But a customer's ability to make this threat credible is restricted because she will find it optimal to continue to hire that expert for problems for which he is most qualified in the face of truthful reports from other experts. My purpose in this article is to clarify the extent to which experts' desire to maintain their "goodwill" reputation works to enhance the quality of their advice in such environments. I also characterize the exact type of consulting behavior that a customer may adopt to achieve the maximum possible quality.

I model the experts' advice (as to who is the right service provider for the problem at hand) as "cheap talk."[2] The advice is given free. The customer's payoff depends solely on whether or not she hires the right expert to treat the problem. I abstract from some issues of general importance but with less direct bearings on the question, such as price competition and moral hazard in service provision. This framework allows me to focus on the core problems of the agents for my main task of delineating the effect of enhancing trust in advice via fostering vested interests in future business. In addition, casual observations suggest that these simplifying assumptions may be appropriate for some cases of interest. In the health service industry, for instance, patients with insurance coverage are not concerned about the cost of individual visits.

In this model, the agents' decision problems are straightforward. In each period, the customer encounters a new problem and decides (i) which and how many (experts) to consult for advice and (ii) which to hire based on the received advice. The expert's decision is, when consulted, whether to mislead the customer to win the current business for short-term gain or to stay trustworthy for future rewards. I analyze an infinitely repeated version of this stage game between a customer and differentiated experts, and characterize "stationary" equilibria.

An intrinsic issue in infinitely repeated games is the credibility of punishment: punishing the cheater typically incurs damage to the punisher as well as the punished, and so is susceptible to renegotiation. To deal with this problem in the context here, I develop a criterion called "recursive credibility" by requiring that the punishment phases be robust to coalitional renegotiation that satisfies an internal consistency condition with respect to its own punishment phases. This criterion effectively selects a unique equilibrium in my model.

My main findings are: (1) Fully honest advice may not be sustained if the profitability of service provision varies widely across problems. (2) As the number of experts increases due to a higher degree of specialization, the maximum equilibrium honesty level deteriorates. (3) Nonetheless, the equilibria that pass the recursive credibility check on their punishment phases implement the same (unique) honesty level regardless of the number of experts. Furthermore, the customer can extract this honesty level by appointing a "panel" of only one or two (but no more) experts and consulting them all the time.

It is not surprising that full honesty is generally not obtained, since the experts would try to cash in on their reputations by misleading the customer if the profitability of the current problem (which is a random draw) is sufficiently high. The other two findings, on the other hand, appear rather counterintuitive at first sight, for one would normally expect that increased competition (be it for service provision or for advice) among a larger number of experts would enhance the quality of service.[3] In the context of my analysis, however, a larger number of potential suppliers does not by itself mean more competition. This is because of asymmetric information, i.e., the service qualities of potential suppliers are not known to the customer before purchase. Instead, the balance between expected gain from cheating and the future income to be forfeited as punishment, as explained below, provides an insight into my results.

As noted above, the consulted panel is dishonest when the profitability is sufficiently high. Since the panel's reports carry no information regarding who is the right specialist in these cases,

---

[2] A message is cheap talk if it is costless (i.e., it does not affect payoffs directly), unverifiable, and nonbinding.

[3] Satterthwaite (1979) shows a result of similar flavor: an increased number of sellers (experts) in a monopolistically competitive market of a "reputation good" may cause the price to rise. Unlike my analysis, central to his result is the search cost that increases as there are more sellers, causing the demand for individual sellers to be less price elastic.

the customer is indifferent as to which expert to hire. If the customer patronizes the panel members in these cases, then their expected future business is enhanced. This means that the experts would suffer a higher loss if dismissed from the panel, so they have more incentive to stay on the panel by being faithful. But if dishonesty is "tolerated" in this sense only for very profitable cases (and thus rather infrequently), the future business will not be a sufficiently large "collateral" to induce honesty in all other cases. As tolerated dishonesty expands to somewhat less-profitable cases, the consulted experts also extend more honesty than before to more-profitable problems. The maximum level of sustainable honesty, therefore, is the highest level of profitability for which the panel members will be honest, given that their dishonesty will be tolerated for more-profitable cases.

The collateral value of each panel member consists of the future business from tolerated dishonesty and the fair share of honest business. As there are more experts due to finer specialization, the latter component gets smaller, hence so does the total collateral value. This accounts for my second result that the maximum sustainable honesty level deteriorates as there are more differentiated experts.

This maximum honesty level is sustained when panel members believe that cheating will forfeit the full collateral value, i.e., the entire future business. This means that, for the sake of maximally punishing the cheater, the customer expects no honesty at all from any expert once cheating is detected (for otherwise the cheater can be truthfully reported as the right specialist, in which case it is subgame optimal for the customer to hire him). The credibility of such a punishment strategy is questionable, though, because if the customer can sustain a certain level of honesty from some panel in equilibrium, there is no reason for her not to be able to induce the same honesty from another panel in the punishment phase. Capturing such internal consistency, the recursive credibility implies that after cheating, the customer would have a new panel of advisors no less honest than the initial, equilibrium panel. This new panel cannot be more honest either, for then the customer would bypass the initial panel.

Since the honesty level remains the same before and after cheating in recursively credible equilibrium as such, every expert including the cheater retains the part of future business associated with truthful reporting no matter what. So this part does not affect the incentive to cheat, hence neither does it affect the sustainable honesty level. Then maximum honesty is obtained when the part of business associated with forgiven dishonesty is most effectively used to raise the collateral value of the panel members. The best way to use it is to concentrate on one or two experts, as explained below. Since this can be done regardless of the total number of experts, the maximum recursively credible honesty level does not vary with it.

Successful cheating requires a majority collusion within the panel, who conspire to mislead the customer into hiring one of them and split the proceeds. The incentive to cheat is lowest when the size of the smallest majority collusion relative to the whole panel is largest, because then each collusion member's share of cheating proceeds would be the smallest relative to his share of collateral business to be forfeited. This happens with one- or two-member panels because the majority collusion and the whole panel coincide for them, establishing the third main finding.

This article contributes to the literature on cheap-talk reputation. Sobel (1985) shows that an "enemy" (an informed agent with interests completely opposed to the decision maker) may build a reputation by mimicking the honest behavior of a "friend" (with interests identical to the decision maker), only to cash it in when the stake is high enough. Benabou and Laroque (1992) generalize Sobel's model by incorporating noisy information in an asset market setting. In a model where an enemy is biased in one direction, Morris (2001) shows that even a friend may have a reputational incentive to lie in the other direction. In these articles, the identity (friend or enemy) of the informed agent is fixed throughout, and therefore reputation building is possible even in a finite horizon. In a model where the identity of the informed agent is drawn independently in each period (as in this article), Kim (1996) shows that infinitely repeated pretrial negotiation can enhance the credibility of cheap talk and improve efficiency. Hermalin (1998) also studies a model with serially uncorrelated identity, but his emphasis is on the effects of repeated cheap-talk communication in reducing the need for other, costly means such as signalling and screening.

Ottaviani and Sorensen (2001) model an expert's reputation differently in that the experts are motivated by exogenous reputational payoff, as in the career concerns literature. A feature that distinguishes the present article from these studies is the direct competition of multiple experts for the customer's trust.

Multiple experts with conflicting interests have been investigated in various static (i.e., one-shot) settings: Gilligan and Krehbiel (1989) and Austen-Smith (1990) in legislative contexts; Shin (1994) in an arbitration with uncertain information partitions and verifiable reports; Lipman and Seppi (1995) in multi-sender/receiver games of sequential talk and partial provability; Krishna and Morgan (2001) in a two-sender/receiver game of sequential talk; and Pesendorfer and Wolinsky (2003) in a credence service context with diagnostic moral hazard, among others. Wolinsky (2002) also studies an environment with multiple experts who share similar preferences but possess different pieces of information. By their one-shot nature, these studies do not consider reputation, which is central in this article.

Another related line of literature is that of the goodwill reputation in quality signalling: firms charging higher prices indeed supply higher-quality experience goods for fear of losing all customers, hence higher future margins, in case they fail to meet the customers' expectations (see Allen, 1984; Klein and Leffler, 1981; Rogerson, 1983; Shapiro, 1983).[4] This literature deals with the issue of overcoming the problem of moral hazard via costly signalling in competitive environments, whereas the current article studies costless communication on the problem of adverse selection in relation-specific situations.[5]

Only a small literature exists on the specific issue addressed in this article, i.e., the incentives to match the needs of a customer with the right specialists among experts of differentiated expertise. Garicano and Santos (2004) is one other article that investigates referrals among differentiated experts, but their focus is on contracting among experts to enhance referral efficiency in a one-shot setting where clients have no strategic role. My focus is on the relation between client and the experts in recurrent relationships.

The rest of the article is organized as follows. In Section 2 I describe the model with two experts. In Section 3 I characterize the category of equilibria in which the customer relies solely on one expert's advice. Section 4 introduces the notion of recursive credibility. In Section 5 I characterize the other category in which the customer invites rivalry by consulting both experts. In Section 6 I extend the analysis to cases with more than two experts. Section 7 contains some concluding remarks. The Appendix contains the proofs, and a companion web Appendix (available at www.rje.org/main/sup-mat.html) demonstrates the robustness of the main results to one realistic extension of the model, i.e., when the customer observes profitability with some noise.

2. Model and preliminaries

■ For the sake of fixing context for mnemonic reasons, I describe the model as a car repair industry in a small town. There are three infinitely lived players, namely, one customer and two mechanics called A and B. The customer experiences exactly one problem with her car in each period $t = 1, 2, \ldots$, which has to be repaired by one of the two mechanics. The type, τ_t, of the problem in period-t is a random variable taking values A or B with even probabilities: mechanic A (B) is better at repairing problems of type A (B). The "lucrativeness" of period t's problem is a random draw θ_t from a common probability distribution function $F(\theta)$ and density $f(\theta)$ defined on $\Re_+$. The expected value of θ, $E(\theta) = \int_0^\infty \theta dF$, is assumed to be finite. The analysis is carried out as if the support of f were unbounded, but the implications for the cases of bounded support are straightforward and are stated in Remark 1 in Section 5.

[4] Hence, like the current article, the notion of reputation in these articles differs from that formalized by Kreps and Wilson (1982a) and Milgrom and Roberts (1982), namely, misleading other players' beliefs on one's *fixed* type by imitating the behavior of another type.

[5] The service in the current article is an experience good because the quality is revealed *ex post*. The role of cheap-talk advising has been explored in the provision of credence services too, e.g., by Pitchik and Schotter (1987). (Credence goods are goods whose quality may never be known to the purchaser.) See also, for example, Wolinsky (1993) and Taylor (1995) for dynamic analysis of credence-goods markets.

The stage game proceeds as follows. When a problem occurs in period t, the customer knows the value of its lucrativeness θ_t, but not its type, τ_t. She consults either mechanic A or B (possibly both) for a diagnosis. Either mechanic correctly identifies the values of τ_t and θ_t when consulted and sends a cheap-talk message regarding the type of problem. Based on the messages received, the customer updates her belief on the problem's type and hires a mechanic to provide a repair service. I use $d_t = A$ (B) to denote that mechanic A (B) is hired.

I set aside the issue of search cost by assuming that there is no cost for either mechanic to identify τ_t and report on it, therefore the consultation is free of charge. I treat consultation activity as private between each mechanic and the customer, so that one mechanic cannot base his report on whether the other had been consulted. This seems realistic, but the imperfect monitoring of the other expert's report incurs some coordination problems in implementing punishment for certain cases, which I discuss in Section 5.

The period payoffs, u for the customer and v_j for mechanic $j(= A, B)$, are functions of τ_t, θ_t and the hiring decision d_t. (There is no price competition as explained in the Introduction.) Letting E_β denote the expectation based on posterior belief β, we assume the following properties:

(U1) Given θ_t, $E_\beta u(\tau_t, \theta_t, d_t = A)$ is bigger than (equal to, smaller than, respectively) $E_\beta u(\tau_t, \theta_t, d_t = B)$ if β puts more (equal, less, respectively) weight on $\tau_t = A$ than B.

(U2) $v_j(\tau_t = A, \theta_t, d_t) = v_j(\tau_t = B, \theta_t, d_t) \geq 0$ is an increasing function of θ_t if $d_t = j$; $v_j(\tau_t, \theta_t, d_t) = 0$ if $d_t \neq j$.

Condition (U1) says that the customer wants to hire the mechanic that she believes is more likely to be the right specialist for the current problem. According to (U2), no matter who the right specialist is, the hired mechanic gets a positive payoff that increases in θ_t, and the other mechanic gets zero.[6] These two conditions are sufficient for all the results in this article. Purely for expositional convenience, however, I fix specific payoff functions in the rest of the article: $u(\tau_t, \theta_t, d_t)$ is a constant $u > 0$ if $\tau_t = d_t$ and is zero otherwise; $v_j(\tau_t, \theta_t, d_t)$ is θ_t if $d_t = j$ and is zero otherwise (this can be obtained by rescaling θ). These payoffs are summarized in Table 1, which lists the payoffs for mechanic A, mechanic B, and the customer in that order.

I examine the situation where this stage game is repeated infinitely and the players discount future payoffs by the same factor, $\delta \in (0, 1)$, and characterize sequential equilibria (naturally extended to infinite games).

Some features of the model are for analytic convenience. The qualitative results of the article remain valid if θ_t is revealed to the customer at the end (rather than at the beginning) of period t, or if the customer observes an imperfect, positively correlated signal of θ_t (see the online Appendix). The same is true when small search costs exist and are reflected in a fixed consultation fee.

The assumption of even prior on the problem's type τ_t, however, is important in my analysis and discussion. For one, the rivalry between the two mechanics would not be on a level playing field if the customer is biased *a priori* toward one of the mechanics. Nonetheless, it does not seem very realistic to assume that the customer truly believes that every sort of problem arises with exactly the same probability. In fact, my analysis covers more plausible environments: sometimes the customer knows the type of problem and needs no consultation, and other times the problem is too complicated or new for the customer to self-diagnose, for which an unbiased prior seems sensible. Formally, my results extend straightforwardly to the cases that τ_t puts uneven probabilities on A and B, while the customer observes a random variable $\hat{\tau}_t$ with noise such that (i) $\hat{\tau}_t = \tau_t$ with a fixed measure, or else (ii) $\hat{\tau}_t$ is uncorrelated with τ_t, and conditional on this event, $\tau_t = A$ and $\tau_t = B$ are equally likely.

No information transmission is possible if the stage game is played once, because either mechanic would claim the problem to be his speciality to try to mislead the customer into hiring

[6] The experts' advice being modelled as cheap talk, legal liability cannot be formally addressed in this article. It is possible, however, to interpret a finite upper bound of f as reflecting the threshold above which cheating is not justified for the expert because it warrants a litigation that results in a compensation proportional to θ_t. My analysis, then, deals with θ_t values for which litigation is not justified for the customer because the expected compensation does not cover the litigation cost.

TABLE 1 **Payoffs for Mechanic A, Mechanic B, and Customer**

	$d_t = A$	$d_t = B$
$\tau_t = A$	$\theta_t, 0, u$	$0, \theta_t, 0$
$\tau_t = B$	$\theta_t, 0, 0$	$0, \theta_t, u$

him.[7] That is, the mechanics *babble*, i.e., send messages that have no correlation with the true type of the problem, and therefore the customer ignores the messages and bases her decision on the prior. In fact, repeating such a babbling equilibrium in every period constitutes an equilibrium of the repeated game, which is a known feature of cheap-talk games.[8]

This article concerns more interesting equilibria of the repeated framework in which effective cheap-talk communication arises by the consideration of reputation. However, fully honest reporting may not be sustained: if the current θ_t is so high that the short-term gain would overcompensate the discounted sum of future losses due to failing the customer's trust, the consulted mechanic has incentives to mislead the customer. Taking such incentives into account, the customer interprets the messages as meaningless. For lower values of θ_t, on the other hand, the potential short-term gain would not justify future losses, so the consulted mechanic reports honestly by sending a particular message if $\tau_t = A$ and another distinct message if $\tau_t = B$. I say that he *recommends* mechanic $j(= A, B)$ if he sends the particular message that he is supposed to send only when $\tau_t = j$.

In light of this observation, it appears most natural for each mechanic to adopt a cutoff strategy in each period t, if consulted: he reports honestly if $\theta_t < \tilde{\theta}_t$ for a certain threshold level $\tilde{\theta}_t$ (the half-open interval $[0, \tilde{\theta}_t)$ is called the *trusted range* for the mechanic), but he babbles if $\theta_t \geq \tilde{\theta}_t$ (the interval $[\tilde{\theta}_t, \infty)$ is called the *distrusted range*). I say that a mechanic *reports with a trust level* $\tilde{\theta}_t$ if he uses this strategy, and I say that he *cheats* if he is supposed to report with a trust level $\tilde{\theta}_t$ but deviates by recommending mechanic j when $\tau_t \neq j$ in the trusted range (i.e., when $\theta_t < \tilde{\theta}_t$).

The core element of reputational consideration in my context is the shift of trust (and future business prospects along with it) anticipated in case of cheating, which feeds back into current incentives to cheat. To focus tightly on this aspect, I investigate "stationary" equilibria in which the mechanics report with the same trust level every period as long as no deviation has taken place. In Section 5, I show that this class of equilibria is sufficient for my purposes. From now on, an equilibrium refers to a stationary equilibrium.

3. Primary-agency equilibrium

■ I say that the customer patronizes a mechanic as a *trusted agent* if, as long as he has not cheated, in each period (i) the customer consults only the trusted agent, (ii) the agent reports with a certain trust level, and (iii) the customer hires the recommended mechanic for repair service in the trusted range and hires the trusted agent in the distrusted range. The equilibria in this section depict situations in which the customer patronizes one of the mechanics as a sole trusted agent, with a view to taming him via enhancing her value as a future customer. On the other hand, she is completely vulnerable to cheating.

A primary-agency equilibrium is described by a sequence of phases, each with a trusted agent. Phase 0, or an initial phase, comprises periods $t = 1, 2, \ldots,$ in which the customer patronizes one of the mechanics as the trusted agent with an initial trust level $\theta^{(0)}$. This trusted agent, whom I label

[7] If the customer can commit to a punishment when the two reports differ (e.g., not hiring either) and the experts cannot collude, then full information can be extracted at zero cost as the mechanism-design literature suggests for correlated types. Such punishment, however, is not subgame perfect for the customer and hence would not be credible in the current context.

[8] See, for example, Blume (1994) and Park (1997) for multiplicity of equilibria in cheap-talk games.

mechanic A, is called the primary agent because the initial phase prevails forever in equilibrium; other phases describe the off-equilibrium paths.

If the primary agent cheats in period t, that is, he recommended the wrong mechanic when $\theta_t < \theta^{(0)}$, the customer finds this out at the end of period t by the realized payoff.[9] Then, phase 1, or a *first backup phase*, starts and prevails in periods $t+1$, $t+2, \ldots,$ in which the customer patronizes the other mechanic (mechanic B) as the trusted agent with a first backup trust level $\theta^{(1)}$. Transition from phase 0 to phase 1 (after such a deviation) would be synchronized by all three players: the customer and the deviator know the deviation and, hence, the transition; the new trusted agent detects the transition when he gets consulted in period $t+1$, and behaves accordingly (see (P2) below).

If the first backup agent cheats in period t' of phase 1, then phase 2, or a *second backup phase*, starts and prevails in periods $t'+1$, $t'+2, \ldots,$ in which the customer switches the trusted agent back to mechanic A but with a trust level $\theta^{(2)}$. Higher-order backup phases, phases $k = 3, 4, \ldots,$ are modelled in an analogous manner, with trust levels $\theta^{(k)}$.

Let $S = \langle \theta^{(0)}, \theta^{(1)}, \ldots \rangle$ denote the players' behavior described above. The behavior of the continuation game at the beginning of phase $k \geq 1$ is denoted by a truncated sequence $S^{(k)} = \langle \theta^{(k)}, \theta^{(k+1)}, \ldots \rangle$ from phase k and onward, with the implicit understanding that mechanic A (B) is the trusted agent in the initial phase of $S^{(k)}$ if k is even (odd). A primary agency equilibrium (hereafter, PAE) is an infinite sequence $S = \langle \theta^{(0)}, \theta^{(1)}, \ldots \rangle$ such that each player's behavior is a best response to those of other players in S and in each $S^{(k)}$ for $k = 1, 2, \ldots.$

To characterize the PAE, consider the trusted agent, say mechanic A, in an arbitrary period t of phase k. He would consider cheating only if $\theta_t < \theta^{(k)}$ and $\tau_t = B$. (Cheating is not feasible if $\theta_t > \theta^{(k)}$ because he is supposed to babble anyway.) Cheating in this case would ensure a short-term gain of θ_t for mechanic A, since the customer follows his advice for sure. On the other hand, cheating will initiate phase $(k+1)$ that will replace phase k in all future periods. Since mechanic A's expected payoff in each future period is

$$\int_{\theta^{(k)}}^{\infty} \theta dF + \frac{1}{2}\int_0^{\theta^{(k)}} \theta dF \quad \text{and} \quad \frac{1}{2}\int_0^{\theta^{(k+1)}} \theta dF \tag{1}$$

in phases k and $(k+1)$, respectively, the punishment for cheating is the discounted sum of the difference. Hence, the equilibrium condition that cheating is not beneficial for all $\theta_t < \theta^{(k)}$ is written as

$$\frac{\delta}{1-\delta}\left(\int_{\theta^{(k)}}^{\infty} \theta dF + \frac{1}{2}\int_0^{\theta^{(k)}} \theta dF - \frac{1}{2}\int_0^{\theta^{(k+1)}} \theta dF\right) - \theta^{(k)} \geq 0. \tag{2}$$

Given $\theta^{(k+1)}$, define $\bar{\theta}(\theta^{(k+1)})$ to be the value of $\theta^{(k)}$ at which (2) is satisfied tightly, i.e., as an equality. Since the left-hand side of (2) decreases in $\theta^{(k)}$, the function $\bar{\theta}(\cdot)$ is well defined and has the property that inequality (2) holds if and only if $\theta^{(k)} \leq \bar{\theta}(\theta^{(k+1)})$. Since the nontrusted agent's behavior is trivially optimal because he does not make any strategic moves, I summarize the optimality of mechanics (experts) in Lemma 1. I state some properties of $\bar{\theta}(\cdot)$ in Lemma 2, which will be used later.

Lemma 1. Each agent's behavior is a best response in a sequence $S = \langle \theta^{(0)}, \theta^{(1)}, \ldots \rangle$ and in each truncated sequence $S^{(k)}$, if and only if

$$0 \leq \theta^{(k)} \leq \bar{\theta}(\theta^{(k+1)}) \qquad \forall\, k = 0, 1, 2, \ldots. \tag{3}$$

Lemma 2. The function $\bar{\theta}(\cdot)$ is a strictly decreasing function. Denoting the unique fixed point of

[9] More generally, there may be some time lag for the customer to discover cheating attempts. This lag exacerbates the incentive to cheat, and hence reduces the honesty level, because the punishment is delayed and more cheating will be possible until the punishment starts. This effect is common across different consulting behavior and hence does not change my main results on comparing them.

$\bar{\theta}(\cdot)$ by θ^*, we have

$$0 < \theta^* = \bar{\theta}(\theta^*) < \bar{\theta}(0) < \frac{\delta}{1-\delta} E(\theta), \tag{4}$$

where $E(\theta) = \int_0^\infty \theta dF$.

Proof. Note that $V(\theta^{(k)}) := \int_{\theta^{(k)}}^\infty \theta dF + (1/2)\int_0^{\theta^{(k)}} \theta dF$ is strictly decreasing in $\theta^{(k)}$ and, hence, so is $W(\theta^{(k)}) := [\delta/(1-\delta)]V(\theta^{(k)}) - \theta^{(k)}$. If $\theta^{(k+1)}$ increases, so must $W(\theta^{(k)})$ to keep (2) satisfied tightly. Therefore, $\bar{\theta}(\cdot)$ is a strictly decreasing function.

From $W(\bar{\theta}(0)) = 0$ and $V(\theta^{(k)}) < E(\theta)$ for all $\theta^{(k)}$, we get the last inequality of (4). Since $W(0) > 0$ and $W(\cdot)$ is a decreasing function, we may deduce $\bar{\theta}(0) > 0$. Finally, note that since the left-hand side of (2) is continuous in $\theta^{(k)}$ and $\theta^{(k+1)}$, so is $\bar{\theta}(\cdot)$. Since $\bar{\theta}(\cdot)$ strictly decreases, there is a unique fixed point θ^* strictly between 0 and $\bar{\theta}(0)$. *Q.E.D.*

Next I check the optimality of the customer's behavior in $S = \langle \theta^{(0)}, \theta^{(1)}, \dots \rangle$ described earlier, which depends on what would happen if the customer were to deviate. Specification of such off-equilibrium paths that supports the customer's behavior as a best response, is not unique. Below I describe one specification that, considering the equilibrium behavior, I believe is sensible. I retain (3) in this discussion. Note that the off-equilibrium behavior I postulate in this and later analyses can be verified in a straightforward way to be compatible with a "consistent assessment" of Kreps and Wilson (1982b). The explanation, however, is lengthy and so is omitted.

When a deviation takes place, the players change their beliefs about the future course of the game. I say that a mechanic assumes a sequence of phases S' actively (passively) in period t if he believes that the initial phase of S' has started in period t with himself (the other mechanic) as the initial trusted agent, to be followed by subsequent phases of S' in cases of cheating. In the special case that S' is the truncated sequence $S^{(k)}$ of the original sequence $S = \langle \theta^{(0)}, \theta^{(1)}, \dots \rangle$, I say that a mechanic assumes phase k in period t (actively if he is the trusted agent in phase k of S, and passively if not).

The customer may deviate from S in either the consultation or the hiring decision. First, I postulate agents' responses to deviations in consultation.

(P1) If the trusted agent, mechanic A, of the initial phase (phase 0) is not consulted in period 1, he assumes phase 1 in period 1. Likewise, if the nontrusted agent, mechanic B, gets consulted in period 1, he assumes phase 1 in period 1.

(P2) Suppose that phase $k(= 0, 1, \dots)$ started in period t. If the trusted agent is not consulted in period $t' > t$ of phase k, he assumes phase $k+1$ in period t'. Likewise, if the nontrusted mechanic gets consulted in period $t' > t$ of phase k, he assumes phase $k+1$ in period t'.

(P3) Suppose the trusted agent, say mechanic A, cheated in period t of phase k. If he is still consulted in period $t+1$, he assumes $\langle \tilde{\theta}^{(k)}, \theta^{(k+1)}, \theta^{(k+2)}, \dots \rangle$ actively in period $t+1$, where $\tilde{\theta}^{(k)} = \min\{\theta^{(k)}, \theta^{(k+1)}\}$.[10] Mechanic B, however, believes that the original phase k (i.e., with the trust level $\theta^{(k)}$) continues to prevail if he is not consulted in period $t+1$.[11]

The other kind of possible deviation by the customer is that she may not follow the trusted agent's recommendation in her hiring decision. I can easily remove incentives for such deviation by postulating the following.

(P4) If either agent detects a deviation in the customer's hiring decision, he attributes it to a simple mistake and does not change his belief about the prevailing phase.

[10] This is as if he believes he has been given a second chance. I take the minimum here not to give the customer an incentive to forgive him. Due to (5), to be derived shortly, this amounts to taking $\tilde{\theta}^{(k)} = \theta^{(k+1)}$. If I postulate $\tilde{\theta}^{(k)} = \theta^{(k)}$ instead, the "recursively credible" equilibrium (to be discussed later) obtains.

[11] He may have suspected cheating by mechanic A because, for instance, he has provided the service when he was not supposed to. However, it is always possible that such experience was due to the customer's deviation in the hiring decision, which does not change agents' beliefs as postulated in (P4).

Note that according to (P2), the customer may deviate in phase k by consulting only the nontrusted agent so as to induce both agents to assume $\langle \theta^{(k+1)}, \theta^{(k+2)}, \dots \rangle$, which would indeed be beneficial for the customer if the trust level is higher in phase $(k+1)$ than in k. Therefore,

$$\theta^{(k)} \geq \theta^{(k+1)} \qquad \forall\, k = 0, 1, 2, \dots, \tag{5}$$

is necessary for the customer not to deviate in any phase of S. Together with (3), this characterizes the PAE as stated in Theorem 1 below, which summarizes the findings so far.

Theorem 1. A sequence $S = \langle \theta^{(0)}, \theta^{(1)}, \dots \rangle$, augmented by the off-equilibrium behavior specified in (P1)–(P4), constitutes a PAE if and only if

$$\theta^{(k+1)} \leq \theta^{(k)} \leq \bar{\theta}(\theta^{(k+1)}) \qquad \forall\, k = 0, 1, 2, \dots . \tag{6}$$

The set of trust levels sustainable in a PAE is $[0, \bar{\theta}(0)]$. The PAE with the maximum trust level is $\langle \bar{\theta}(0), 0, 0, \dots \rangle$.

Proof. I already showed the optimality of the experts in Lemma 1. Below I show the optimality of the customer's behavior under (P1)–(P4) and (6), for which it suffices to verify that she cannot expect better advice from any expert after some deviation than the current trusted agent. I do this for $S = \langle \theta^{(0)}, \theta^{(1)}, \dots \rangle$. The same argument applies to any truncated sequence $S^{(k)}$.

Consider an arbitrary sequence of deviations by the customer, either in the consultation decision or the hiring decision. First, note that any deviation in the hiring decision does not affect the reporting strategy of either expert by (P4). In addition, by (P4), if the customer makes multiple deviations in consultation over time, each agent updates his belief on the prevailing phase according to the relevant postulate among (P1)–(P3) at each incidence of deviation. It should be noted that agents may not have synchronized beliefs on the off-equilibrium paths because they may diverge in detecting deviations.[12] Nonetheless, for each agent in each period, the continuation game is equivalent to the start of the initial phase of a sequence that satisfies (3). This is obvious because they believe themselves to be in some phase of the original sequence S except for mechanic A described in (P3), in which case (3) follows because $\tilde{\theta}^{(k)} \leq \theta^{(k)} \leq \bar{\theta}(\theta^{(k+1)})$. (Hence, by Lemma 1, each agent's behavior postulated above is a best response given his belief at that point in time.) In any period after any sequence of deviations by the customer, therefore, either agent would report with a trust level not exceeding the initial equilibrium trust level by (6). This establishes the optimality of the customer's behavior in S.

As for the latter two assertions of the theorem, note that the initial trust level prevails forever. Because $\bar{\theta}(\cdot)$ is a decreasing function, the range of $\theta^{(0)}$ in (6) is largest when $\theta^{(1)} = 0$. Hence, the set of trust levels sustainable by a PAE is $[0, \bar{\theta}(0)]$. Clearly, the PAE with the maximum trust level is $\langle \bar{\theta}(0), 0, 0, \dots \rangle$. *Q.E.D.*

4. Recursive credibility

■ I believe, however, that the equilibrium condition (6) leaves too much freedom in specifying the backup trust levels. In particular, the maximum trust level $\bar{\theta}(0)$ discussed above is supported by the extreme backup trust levels $\theta^{(k)} = 0$ for all $k = 1, 2, \dots$, i.e., by the threat that the primary agent will never be hired again if he ever cheats. We doubt that such a threat is really credible: once the first backup phase starts, the nontrusted agent may approach the customer and offer a "coalitional deviation" to start another PAE with a higher trust level, which would be beneficial for both the customer and himself. It is also conceivable that the customer may initiate such offers. The same argument applies to higher-order backup phases.

But not every such deviation would be viable. Specifically, a deviation would not be viable

[12] For example, if both agents are consulted in phase 0, mechanic A assumes phase 0 in the next period and mechanic B assumes phase 1. Each agent's strategy based on his own updating of the phase is optimal despite such discrepancies, either because he is unaware of them according to a consistent belief profile or because they do not affect his incentives.

if it were itself to be overturned by another deviation. For such coalitional deviations in backup phases to be valid, therefore, the new equilibria to be adopted by the deviations need to be robust to the same kind of credibility check. That is, internal consistency requires that the validity of deviations be judged by the same criterion used to judge the original equilibrium. This makes the concept of credibility (yet to be defined) recursive.

My notion of credibility is a variant of the coalition-proofness of Bernheim, Peleg, and Whinston (1987) and its extension by Ferreira (1996). These concepts are also recursive, but they are developed for cases with finite recursion. In my environment, the recursion is inherently infinite and the definition is circular. Nonetheless, it allows me to identify the unique PAE that conforms to the definition.

Definition 1. A PAE *overrides* another PAE if the initial trust level of the former is strictly bigger than that of the latter.

(i) A PAE $S = \langle \theta^{(0)}, \theta^{(1)}, \dots \rangle$ is *round-1 credible* if there does not exist a round-1 credible PAE that overrides the truncation $S^{(1)} = \langle \theta^{(1)}, \theta^{(2)}, \dots \rangle$.

(ii) Let $k > 1$ and assume that a round-k' credible PAE has been defined for all $k' < k$. Then, a PAE S is *round-k credible* if

(a) $S^{(1)}$ is round-$(k-1)$ credible, and

(b) there does not exist a round-k credible PAE that overrides $S^{(1)}$.

(iii) A PAE S is *recursively credible* if it is round-k credible for all $k = 1, 2, \dots$.

This definition implies the desired property that a recursively credible PAE is backed up by a sequence of punishment phases that is also recursively credible and is not to be overturned by a deviation that passes the same credibility check.

But due to the circularity of the definition, I cannot check the credibility of an individual PAE separately: the round-k credibility of a PAE depends on that of other PAE's, and vice versa. Instead, I need to find the sets of round-k credible PAE's, inductively on k, and then take the intersection to obtain the set of recursively credible PAE's. Rather than going through the full process,[13] I take a shortcut to identify a recursively credible PAE that turns out to be the unique one.

A round-k credible PAE $S = \langle \theta^{(0)}, \theta^{(1)}, \dots \rangle$ cannot have $\theta^{(0)} > \theta^{(1)}$, because if so, S itself overrides $S^{(1)}$, contradicting condition (b) of part (ii) above. Together with condition (5) of PAE, it follows that $\theta^{(0)} = \theta^{(1)}$. Since this holds for every k and every truncation of a recursively credible PAE is also recursively credible by definition, any recursively credible PAE must have the same trust level sustained in all phases.

Recall that the sustainable honesty level decreases in the backup trust level, i.e., $\bar{\theta}(\cdot)$ is a decreasing function. Hence, the maximum (constant) trust level sustainable in all phases is the fixed point θ^* of $\bar{\theta}(\cdot)$. Although any lower trust level is also sustainable in all phases, such a PAE is overriden by one that has θ^* as the constant trust level and, therefore, is not recursively credible. Indeed, we have the following.

Theorem 2. $S^* = \langle \theta^*, \theta^*, \dots \rangle$ is the unique recursively credible PAE.

Proof. As discussed above, the first two trust levels of a round-k credible PAE must be the same number between zero and θ^*.

Consider $S^* = \langle \theta^*, \theta^*, \dots \rangle$. Since θ^* is the maximum initial trust level for round-1 credible PAE's, no round-1 credible PAE overrides the first truncation of S^* (which coincides with S^*). Hence, S^* is round-1 credible.

[13] Briefly, the set RC(1) of round-1 credible PAE's consists of those with $\theta^{(0)} = \theta^{(1)}$ as explained in the next paragraph. Note that $\theta^{(0)} = \theta^{(1)} \leq \theta^*$ by (6). The inequality cannot be strict, otherwise it would be overriden by $S^* = \langle \theta^*, \theta^*, \cdots \rangle$. Hence, RC(1) consists of PAE's with $\theta^{(0)} = \theta^{(1)} = \theta^*$. Inductively, RC($k$) consists of the ones with $\theta^{(k')} = \theta^*$ for $0 \leq k' \leq k$.

Next, let $k > 1$ and suppose S^* is round-$(k-1)$ credible. Then, condition (a) of part (ii) above is trivial. By an argument analogous to the one in the previous paragraph, condition (b) of part (ii) is also satisfied, and therefore S^* is round-k credible. Therefore, S^* is recursively credible.

Finally, any constant sequence $S' = \langle \theta', \theta', \dots \rangle$ with $\theta' < \theta^*$ is clearly overriden by S^* and, hence, is not round-k credible for any k. This proves the uniqueness. *Q.E.D.*

5. Rivalry-agency equilibrium

■ Rivalry-agency equilibria depict situations in which the customer holds the two mechanics in check and invites rivalry by consulting them both. Relative to PAE, she is less vulnerable to cheating attempts because she receives two independent opinions. At the same time, she is valued less by each mechanic because she is "shared" between the two.

Formally, an initial phase comprises periods $t = 1, 2, \dots,$ in which mechanics A and B report with initial trust levels θ_A and θ_B, respectively, where I assume $\theta_A \geq \theta_B$ without loss of generality, and the customer responds as follows: (i) if $\theta_t \geq \theta_A$, she hires mechanics A and B with probabilities p and $1-p$, respectively; (ii) if $\theta_A > \theta_t \geq \theta_B$, she hires the mechanic that mechanic A recommends; and (iii) if $\theta_B > \theta_t$, she hires the recommended mechanic if the recommendations coincide, but if they do not coincide she hires mechanics A and B with probabilities q and $1-q$, respectively.[14]

If one of the mechanics, say mechanic A, cheats in period t, the customer identifies the cheater at the end of period t, and a first backup phase (phase 1) prevails in periods $t+1, t+2, \dots,$ in which the customer patronizes mechanic B as the trusted agent who reports with a first backup trust level $\theta^{(1)}$, i.e., in the same manner as in a PAE explained in Section 3. Higher-order backup phases are modelled in the same manner.

Transition to the first backup phase needs some further explanation because it may not be synchronized among all three players due to imperfect monitoring. For example, suppose that mechanic A cheated in period t of the initial phase but the customer hired the right agent, mechanic B, as a result of randomization. Since mechanic B did not observe mechanic A's report, he would not have detected any deviation. Therefore, he would still report with trust level θ_B in period $t+1$, when he should report with $\theta^{(1)}$.

For expositional convenience, I separate the effects of inviting rivalry from those of imperfect monitoring (which is not a central concern of this article) by postulating the following:

(P5) Reports of each mechanic are retained as indisputable evidence. The mechanics may request these (written) reports. The customer may provide them upon such requests, or voluntarily, or withhold them, at the end of each period.[15]

In the remainder of this section I assume that the agents request the other mechanic's report in each period of the initial phase, so as to detect any deviation right away and to become the sole trusted mechanic, which is potentially profitable. In addition, each mechanic takes the customer's refusal to provide the other mechanic's report as evidence of cheating by him. Then, the transition to the first backup phase is unambiguously coordinated by all three players. I stress here that the effects of imperfect monitoring, if taken into account by removing (P5), do not change my main results and insights mentioned in the Introduction. Remark 2 at the end of this section elaborates on this point.

I denote the players' behavior in successive phases described above by a modified sequence $S_r = \langle (\theta_A, \theta_B), \theta^{(1)}, \theta^{(2)}, \dots \rangle$, which I refer to as a *rivalry sequence*. A rivalry sequence S_r is a *rivalry agency equilibrium* (hereafter, RAE) if each player's behavior is a best response to those of other players in S_r and in each $S_r^{(k)}$ for $k = 1, 2, \dots$. Since the initial phase prevails forever, the

[14] Note that p and q can be functions of θ_t.

[15] This is weaker than assuming observability of the report because they can be withheld. The reports are still cheap-talk messages and cannot be used effectively in a contract: the two possible reports ("$\tau_t = A$" and "$\tau_t = B$") are distinguished only by their truthfulness, which is not observable by a third party.

effective trust level of an RAE is θ_A, the higher of the two initial trust levels. An RAE is *symmetric* if $\theta_A = \theta_B$ and $p = q = 1/2$.

By definition, the backup-phase truncation, $S_r^{(1)} = \langle \theta^{(1)}, \theta^{(2)}, \ldots \rangle$, of an RAE S_r constitutes a PAE described in Theorem 1. Hence, in the remainder I make it a custom that the backup-phase truncation of a rivalry sequence is a PAE. In particular, I take (6) for granted for $k = 1, 2, \ldots$. Then, the best-response property is automatic in every backup phase. Below I focus on the initial phase of S_r.

First, I show in the next lemma that the effective trust level of any RAE is implementable by a symmetric RAE. The basic intuition is that (i) pushing θ_A above θ_B does not help in enhancing the effective trust level because, when $\theta_t \in [\theta_B, \theta_A]$, mechanic A's cheating attempt is assured of success (unlike for $\theta_t < \theta_B$, in which case he can succeed with a 50% chance) and hence he would have a greater incentive to cheat, and (ii) given $\theta_A = \theta_B$, unequal treatment (i.e., $p \neq 1/2$ or $q \neq 1/2$) would increase the incentive to cheat for the less favorably treated mechanic and, consequently, lower the effective trust level.

Lemma 3. Suppose that each agent's behavior is a best response in a rivalry sequence $\langle(\theta_A, \theta_B), \theta^{(1)}, \theta^{(2)}, \ldots\rangle$ for some values of p and q. Then, so it is in a symmetric rivalry sequence $\langle(\theta_A, \tilde{\theta}_B = \theta_A), \theta^{(1)}, \theta^{(2)}, \ldots\rangle$ for $p = q = 1/2$.

Proof. See the Appendix.

In light of Lemma 3, I focus on symmetric RAE from now on and denote the common initial trust level by $\theta^{(0)}$. To derive equilibrium conditions, consider either agent, say mechanic A, who would consider cheating in the initial phase, i.e., when $\theta_t < \theta^{(0)}$ and $\tau_t = B$. Compared with the case when he is the sole trusted agent (which has been analyzed in Section 3), there are two differences: (i) the probability of success is only 1/2 if he cheats, and (ii) he gets to provide the service with a probability 1/2 when $\theta_{t'} \geq \theta^{(0)}$ in each future period t' if he does not cheat. The best-response condition, therefore, is a variant of the inequality (2) that accommodates these two differences: neither agent would ever have an incentive to cheat in the initial period if and only if

$$\frac{\delta}{1-\delta}\left(\frac{1}{2}\int_0^{\infty} \theta dF - \frac{1}{2}\int_0^{\theta^{(1)}} \theta dF\right) - \frac{1}{2}\theta^{(0)} \geq 0. \tag{7}$$

Given $\theta^{(1)}$, define $\bar{\theta}_r(\theta^{(1)})$ to be the value of $\theta^{(0)}$ at which (7) is satisfied tightly:

$$\bar{\theta}_r(\theta^{(1)}) = \frac{\delta}{1-\delta}\left(E(\theta) - \int_0^{\theta^{(1)}} \theta dF\right). \tag{8}$$

Then, $\bar{\theta}_r(\cdot)$ is a well-defined, decreasing function with the property that (7) holds if and only if $\theta^{(0)} \leq \bar{\theta}_r(\theta^{(1)})$. Furthermore, combining with (4), we have

$$\bar{\theta}_r(\theta^*) = \theta^* = \bar{\theta}(\theta^*) < \bar{\theta}(0) < \frac{\delta}{1-\delta}E(\theta) = \bar{\theta}_r(0), \tag{9}$$

where the first two equalities follow because (2) and (7) are equivalent when $\theta^{(1)} = \theta^{(0)}$.

Analogously to before, I characterize the RAE by ensuring that the experts will not cheat given the backup trust level and that the customer has no incentive to deviate both in consultation and hiring decisions. To do this, I modify (P1) and (P3) to (P1a) and (P3a) below, respectively, to accommodate the initial phase of an RAE.

(P1a) If an agent is not consulted in period t of phase 0, he assumes $\langle \theta^{(1)}, \theta^{(2)}, \ldots \rangle$ passively in period t. If an agent is not allowed to see the other mechanic's report in period t, he assumes $\langle \theta^{(1)}, \theta^{(2)}, \ldots \rangle$ actively in period $t + 1$.

(P3a) Suppose that an agent, say mechanic A, cheated in period t of phase 0. If he is still consulted in period $t+1$, he assumes $\langle \theta^{(1)}, \theta^{(2)}, \dots \rangle$ actively in period $t+1$. If the other mechanic, B, is not consulted in period $t+1$, he assumes $\langle \theta^{(1)}, \theta^{(2)}, \dots \rangle$ passively in period $t+1$.

Theorem 3. A rivalry sequence $\langle (\theta^{(0)}, \theta^{(0)}), \theta^{(1)}, \dots \rangle$, augmented by the off-equilibrium behavior as specified in (P1a), (P2), (P3), (P3b), and (P4), constitutes an RAE if and only if

$$\theta^{(1)} \leq \theta^{(0)} \leq \bar{\theta}_r(\theta^{(1)}) \quad \text{and} \quad \theta^{(k+1)} \leq \theta^{(k)} \leq \bar{\theta}(\theta^{(k+1)}) \qquad \forall\, k \geq 1. \tag{10}$$

The set of trust levels sustainable in an RAE is $[0, (\delta/(1-\delta))E(\theta)]$.

Proof. See the Appendix.

Now I apply the credibility argument of the backup phases that has been developed in the previous section. In the same spirit, I define an RAE S_r to be *recursively credible* if the truncation $S_r^{(1)}$ is a recursively credible PAE and there does not exist a recursively credible PAE or RAE that overrides $S_r^{(1)}$. Since $S^* = \langle \theta^*, \theta^*, \dots \rangle$ is the unique recursively credible PAE, any recursively credible RAE S_r should have $S_r^{(1)} = S^*$. Since θ^* is the fixed point of $\bar{\theta}_r(\cdot)$, S_r cannot have an initial trust level higher than θ^* by (10), and hence cannot override $S_r^{(1)} = S^*$. It immediately follows that $S_r^* = \langle (\theta^*, \theta^*), \theta^*, \theta^*, \dots \rangle$ is the unique RAE that is recursively credible.

Theorem 4. The sequence $S_r^* = \langle (\theta^*, \theta^*), \theta^*, \theta^*, \dots \rangle$ is the unique recursively credible RAE.

It is worth noting from (9) that the maximum effective trust level, $\bar{\theta}_r(0) = [\delta/(1-\delta)]E(\theta)$, of an RAE is higher than that of a PAE, $\bar{\theta}(0)$. It turns out that $\bar{\theta}_r(0)$ is indeed the absolute upper bound of θ_t for which a mechanic may report truthfully in *any* stationary equilibrium, including the ones in which mechanics may be consulted randomly and/or they may not report with a trust level. The result is stated in Lemma 4 and is proved in the Appendix. We may say that an equilibrium exhibits truthful revelation for a certain level θ' if the customer hires the right specialist whenever $\theta_t = \theta'$.

Lemma 4. If an equilibrium exhibits truthful revelation for θ', then $\theta' \leq [\delta/(1-\delta)]E(\theta)$.

This result justifies my practice in this article of focusing on equilibria in which players' continuation strategies are independent of past history other than deviations, because no other stationary equilibrium exists with a higher level of honesty. Furthermore, the equilibria considered in this and previous sections effectively cover all equilibria for my purposes, in the sense that for any equilibrium, there exists a PAE or an RAE with the same consumer's and mechanic's surpluses.[16]

Remark 1. Up to now the results have been stated presuming that θ can take any positive value, i.e., the support of f is unbounded. If θ is bounded by $\theta^{\max} < \infty$, the maximum equilibrium trust levels ($\bar{\theta}(0)$ for PAE and $\bar{\theta}_r(0)$ for RAE) may exceed $\theta^{\max}$, in which case complete honesty is sustainable in equilibrium. Nonetheless, complete honesty can never be supported in recursively credible equilibria: since recursive credibility requires identical trust levels in every phase, $\theta^{(0)} = \theta^{(1)} = \theta^{\max}$ cannot be sustained, because if so, cheating clearly pays off in phase 0. This result also applies to the cases of more than two experts in the next section.

Remark 2. If (P5) is removed, imperfect monitoring distills the effects of punishment in an RAE: in the case that a cheating attempt is not detected by the other mechanic, the customer would keep quiet about it and let the innocent mechanic continue to report with the initial trust level $\theta^{(0)}$ rather than triggering a lower, backup trust level $\theta^{(1)}$. The distilling effect is greatest if the random

[16] To see this, observe that in each period t along an arbitrary equilibrium, the set of θ_t for truthful revelation is a subset of $[0, \bar{\theta}_r(0)]$ by Lemma 4. Therefore, the expected consumer's surplus is lower than that in the "most honest" RAE, $\langle (\bar{\theta}_r(0), \bar{\theta}_r(0)), 0, 0, \dots \rangle$. The total expected mechanic's surplus is the same in all equilibria at $[1/(1-\delta)]E(\theta)$. Therefore, the social surplus is higher in the most honest RAE than in the one arbitrarily chosen above. Hence, one can find an RAE with a lower initial trust level that replicates the same total social surplus as the latter.

hiring decision in the distrusted range is made in public, such as alternation, so that the customer may not punish the cheater in a disguised manner (e.g., hire the cheater less than proportionately as if by chance) when cheating is not detected by the other mechanic. It turns out that if public randomization is used, $\bar{\theta}_r(\theta^{(1)})$ coincides with $\bar{\theta}(\theta^{(1)})$ derived in Section 3 for PAE. Otherwise, the equilibria are much more complex,[17] but the overall effect of imperfect monitoring is still to alleviate the punishment and reduce the sustainable trust level (by a lesser degree than the public randomization). This is indirectly verified by the fact that Lemma 4 is proved without (P5). Hence, the customer would commit to (P5) if possible. Finally and importantly, this effect is absent if $\theta^{(0)} = \theta^{(1)}$, because then the customer has no reason to try to stay in the initial phase after a cheating attempt. So, (P5) has no effect on recursively credible equilibria.

6. Extension to more experts

■ I extend the analysis to cases in which there are more than two types of problems and there is one expert for each type of problem. In each period t the consulted agent(s) reports after accurately learning the values θ_t and $\tau_t (= A, B, \ldots, N)$. There being a larger number of experts due to finer differentiation, the degree of rivalry among them is potentially higher. My main concern is its effects on the sustainable trust level.

□ **Primary-agency equilibrium with N agents.** The concept of PAE naturally extends to $N\ (\geq 2)$ experts: in each phase k a trusted agent reports with a trust level $\theta^{(k)}$ and a deviation by the trusted agent would initiate phase $(k+1)$, in which the customer adopts another agent as a new trusted agent who reports with a trust level $\theta^{(k+1)}$. (The exact sequence of trusted agents in successive phases does not matter as long as the trusted agents are different in any two consecutive phases.) As before, I denote such successive phases by a sequence $\langle \theta^{(0)}, \theta^{(1)}, \ldots \rangle$. The only difference from the two-experts case is that now each agent gets to provide the service with probability $1/N$ in each period if θ_t falls in the trusted range. The best-response condition for agents, therefore, is a variant of the inequality (2) that reflects this difference: each agent's behavior is a best response in phase k if and only if

$$\frac{\delta}{1-\delta}\left(\int_{\theta^{(k)}}^{\infty} \theta dF + \frac{1}{N}\int_0^{\theta^{(k)}} \theta dF - \frac{1}{N}\int_0^{\theta^{(k+1)}} \theta dF\right) - \theta^{(k)} \geq 0. \tag{11}$$

Defining $\bar{\theta}^N(\theta^{(k+1)})$ to be the value of $\theta^{(k)}$ at which (11) is satisfied tightly, the agents' behavior is optimal if and only if $\theta^{(k)} \leq \bar{\theta}^N(\theta^{(k+1)})$, $k = 0, 1, 2, \ldots$. In addition, I need $\theta^{(k+1)} \leq \theta^{(k)}$ for all k, because otherwise the customer would maneuver a transition from phase k to $(k+1)$. Hence, a PAE is characterized by[18]

$$\theta^{(k+1)} \leq \theta^{(k)} \leq \bar{\theta}^N(\theta^{(k+1)}) \qquad \forall\, k = 0, 1, 2, \ldots. \tag{12}$$

As before, $\bar{\theta}^N(\cdot)$ is a well-defined, strictly decreasing function. The unique fixed point of $\bar{\theta}^N(\cdot)$, denoted by θ^*, is independent of N, as is evident from (11): the two terms with coefficient $1/N$ cancel each other out when $\theta^{(k)} = \theta^{(k+1)}$. It follows from (12) that the range of possible initial trust level is $[0, \bar{\theta}^N(0)]$, and the range of possible backup trust level is $[0, \theta^*]$.

It is straightforward to verify that $\bar{\theta}^N(\theta) > \bar{\theta}^{N+1}(\theta)$ for $\theta < \theta^*$. The intuition is as follows. Since each agent provides the service less frequently in the trusted range for a larger N, the expected future payoff after cheating is lower, which discourages cheating; by the same reasoning, the expected future payoff from staying faithful is also lower, which encourages cheating. If

[17] The customer may punish the cheater in a disguised manner, and the innocent mechanic tries to take this possibility into account in his response, which in turn feeds back into the degree of disguised punishment by the customer.

[18] A specification of the off-equilibrium paths for optimality of the customer's behavior is a straightforward modification of the postulates discussed in Section 3 and hence is omitted.

$\theta < \theta^*$, the backup trust level is low enough for the latter effect to dominate the former. (They exactly cancel out if $\theta = \theta^*$, hence the same fixed point for all $\bar{\theta}^N(\cdot)$ as noted above.) So, we have $\bar{\theta}^2(0) > \bar{\theta}^3(0) > \cdots$, that is, the honesty level that a PAE can sustain deteriorates as there are more experts.

I now impose the credibility criterion on backup phases. The definition of recursively credible PAE introduced in Section 3 applies to N-experts cases, too. Furthermore, by exactly the same argument as before, it is easy to show that the sequence $S^* = \langle \theta^*, \theta^*, \dots \rangle$ is the unique PAE that is recursively credible, regardless of the number of experts. The findings are summarized below.

Theorem 5. Suppose there are N experts. A sequence $\langle \theta^{(0)}, \theta^{(1)}, \dots \rangle$ constitutes a PAE if and only if (12) holds. The set of trust levels sustainable in a PAE is $[0, \bar{\theta}^N(0)]$, where $\bar{\theta}^N(0)$ decreases in N. The sequence $S^* = \langle \theta^*, \theta^*, \dots \rangle$ is the unique recursively credible PAE for each $N = 2, 3, \dots$.

Corollary 1. As there are more experts, (i) the maximum honesty level sustainable by a PAE decreases, but (ii) the honesty level of recursively credible PAE stays the same at θ^*.

□ **Collusion and *n*-rivalry-agency equilibrium.** In the initial phase of an n-rivalry-agency equilibrium (hereafter, n-RAE) the customer bases her decision on the reports of n agents, $2 \leq n \leq N$, each of whom reports with his own trust level. For a consistent comparison, I focus on "symmetric" equilibria in which the n agents report with a common, initial trust level $\theta^{(0)}$; the customer then hires one of the n agents for repair service with even probability, $1/n$, if $\theta_t \geq \theta^{(0)}$, and hires the most recommended agent if $\theta_t < \theta^{(0)}$ (if there is a tie, she evenly randomizes between the most recommended mechanics). Such a group of n agents is referred to as a *panel*. We may say that the customer *trusts* the panel if she behaves as described above.

If an agent in the panel deviates by cheating, the customer would punish him by not consulting him in the future. But there exists some uncertainty about what kind of backup phase she will resort to. For example, she may keep all noncheaters (i.e, $(n - 1)$-RAE), or she may lose interest in RAE's altogether and resort to a PAE. However, what determines the incentives to cheat, and consequently the sustainable honesty level, is the backup trust level (not the exact form of backup panel) that will prevail in the punishment phase.[19]

In line with the previous sections, I first find the maximum level of honesty sustainable by an n-RAE without restrictions on the backup phases: since lower backup trust level induces higher initial trust level, I do this by setting all backup trust levels at zero. Then I impose the recursive-credibility criterion.

Consider a panel member in the initial phase of an n-RAE. If $n \geq 3$, unilateral cheating is never profitable because it would not change the customer's hiring decision (because all other panel members report honestly) but would initiate the backup phase. Since this is true for all θ_t in the trusted range regardless of the value of $\theta^{(0)}$, full honesty would be sustainable if only unilateral deviations were feasible.

In the considered environment, however, collusive deviations arise as a relevant issue both theoretically and practically. For example, with three experts A, B, and C, it certainly seems possible that agents B and C agree to report B when $\tau_t = A$ and split the proceeds. Hence, I consider collusion by agents who may conspire to misreport in a coordinated way to mislead the customer's decision and to split the proceeds evenly among themselves.[20]

To determine the honesty level sustainable by an n-RAE, I need to consider the collusion where there is the most incentive to cheat. When n is an odd number, this collusion is clearly one consisting of $(n + 1)/2$ members: as this is the smallest number of members whose cheating

[19] This is no longer true if the customer observes the value of θ_t imperfectly, because then punishment phases are bound to take place along the equilibrium due to the customer's misperception of θ_t, so details of future panels matter for current incentives. Nevertheless, the core relationship between the panel size and the incentive to cheat remains and so do the main results, as explained in detail in the online Appendix.

[20] Because of symmetry, collusion is easier to form when the proceeds are split evenly. However, I do not discuss the issue of enforceability of collusive agreement, which is beyond the scope of this article.

attempt would succeed, each collusion member's share of the proceeds would be largest.[21] When n is even, we need to compare a collusion with $n/2$ members and one with one more member: each member's share would be larger in the former, but their cheating attempt would succeed only half of the time. Because the halved success probability has the dominant effect in the former, the latter collusion is the one with the most incentive to cheat in a panel of n members. This latter collusion is also the collusion where there is the most incentive to cheat in a panel of $n+1$ members as well ($n+1$ is odd). Note that in this collusion, members are less inclined to cheat in a panel of n than in one of $n+1$, because each member has more to lose when the future business of the panel is to be shared among fewer members. This means that a higher honesty level can be sustained by an even n-RAE than by an odd $(n+1)$-RAE. Furthermore, the honesty level that a given n-RAE can sustain is lower as there are more experts due to finer specialization, because then each expert has a smaller share of future business and, therefore, has less to lose. Summarizing, as there are more experts due to finer differentiation, the maximum trust level sustainable by any n-RAE ($2 \leq n \leq N$) strictly deteriorates.

With a given number of experts, N, the arguments above establish that the panel that sustains the maximum honesty consists of an even number of members, but this is not very useful as a practical guide because the exact panel size is not pinned down. Furthermore, the viability of this panel is questionable due to its vulnerability to renegotiations. By imposing recursive credibility, I derive a clearer guideline in this regard.

I generalize Definition 1 in Section 4 to cover all the cases and equilibria considered in this article. Specifically, for N and $n = 1, 2, \ldots, N$, an n-RAE is an infinite sequence $S_{n/N}$ of phases, each phase characterized by a panel (contingent on the identity of cheaters in the previous phase) and the associated trust level (common for all panel members), such that (i) the initial panel size is n, and (ii) each player's behavior in each phase k as described earlier is a best response in the truncation $S^{(k)}_{n/N}$. (Here, a one-member panel is a trusted agent and $S^{(0)}_{n/N} = S_{n/N}$.) By definition, therefore, a truncation $S^{(k)}_{n/N}$ is an m-RAE where m is a number between 1 and N: in particular, $m > n$ is possible. However, the cheaters in one phase are not included in the panel of the next phase, because the customer extracts a higher trust level in this way. For $N > 2$, the set of 1-RAE's includes all PAE's and more.

In earlier discussions on the optimality of the agents, I assumed that the customer would patronize a trusted agent in each backup phase. In "more general" n-RAE's described in the previous paragraph, the backup phases may be served by a panel. As argued earlier, however, what matters is the trust level of the subsequent backup panel, not its size. Therefore, earlier characterizations of equilibria, such as lemmas and theorems (except for the uniqueness of the recursively credible equilibrium), are valid for the more general n-RAE's. In particular, the functions $\bar{\theta}(\cdot)$, $\bar{\theta}_r(\cdot)$, and $\bar{\theta}^N_n$ are valid.

Definition 2. An m-RAE *overrides* an n-RAE if the initial trust level of the former is strictly bigger than that of the latter.

(i) An n-RAE $S_{n/N}$ is *round-1 credible* if there does not exist a round-1 credible m-RAE that overrides the truncation $S^{(1)}_{n/N}$.

(ii) Let $k > 1$ and assume that a round-k' credible n-RAE has been defined for all $k' < k$ and all $n = 1, \ldots, N$. Then, an n-RAE $S_{n/N}$ is *round-k credible* if

(a) $S^{(1)}_{n/N}$ is round-$(k-1)$ credible, and

(b) there does not exist a round-k credible m-RAE that overrides $S^{(1)}_{n/N}$.

(iii) An n-RAE $S_{n/N}$ is *recursively credible* if it is round-k credible for all $k = 1, 2, \ldots$.

[21] I consider deviations in which the collusion members recommend one of them, because recommending another mechanic outside the panel is feasible but would not be as profitable, since this mechanic would demand his share, too, when he performs the service.

This is a straightforward generalization of Definition 1 in Section 4. Consequently, an argument exactly analogous to the one in Section 4 allows me to deduce that candidates for a recursively credible n-RAE have the same trust levels, between 0 and θ^*, for all phases.

By the same logic as used for the case $N = 2$ earlier, one can verify that an n-RAE $S_{n/N}$ is recursively credible if and only if θ^* is the common trust level for all phases. I have already found such equilibria for $n = 1$ and 2: $S^* = \langle \theta^*, \theta^*, \dots \rangle$ is a recursively credible PAE and $S_r^* = \langle (\theta^*, \theta^*), \theta^*, \dots \rangle$ is a recursively credible 2-RAE.

However, it turns out that there is no n-RAE that is recursively credible for $n \geq 3$. In particular, given θ^* as the backup trust level, it is not possible to support θ^* as the initial trust level if $n \geq 3$. The basic reason is that the panel members have more incentive to cheat because the proceeds are shared within a smaller group than the future business from forgiven dishonesty will be when staying faithful. I formalize the discussion on n-RAE's below.

Theorem 6. Suppose there are $N \geq 3$ differentiated experts. The maximum trust level sustainable by any n-RAE $S_{n/N}$, $2 \leq n \leq N$, strictly decreases in N. Recursively credible n-RAE's exist for $n = 1, 2$: they have the same trust level, θ^*, for all phases, and in each phase the panel has one or two members. For $n \geq 3$, a recursively credible n-RAE does not exist.

Proof. See the Appendix.

7. Concluding remarks

■ I have studied how reputational considerations may enhance cheap-talk communication in an environment of recurrent relationships between a customer and multiple experts who provide professional services with differentiated specialties. The results on recursively credible equilibria suggest that the customer may extract more reliable advice by building an exclusive consultation relationship with one or two experts, because it has the effect of increasing the vested interest between them. However, she cannot induce completely honest advice from them.

The key characteristics of the model are present in diverse environments of professional services, although the details vary. The health care market appears particularly interesting because a primary care system and a self-referral system coexist: these are analogues of the two categories of consultation behavior in my analysis. Large companies trust their in-house attorneys or retainer law firms to decide optimally whether to seek more specialized legal service from outside. Financial consultants both advise on the right types of investment for their clients' individual situations and serve as a broker. Other examples include professional repair service providers such as auto mechanics. Generally, it seems difficult to enforce quality in professional service industries from outside, because the quality of service may be subject to professional judgment and may not be unequivocally measured. Therefore, it is worth asking how and what level of quality can be sustained endogenously within the market. These are the questions on which this article tries to shed some light.

Appendix

■ Proofs of Lemmas 3 and 4 and Theorems 3 and 6 follow.

Proof of Lemma 3. First I find mechanic B's optimality condition in the initial phase of a rivalry sequence $\langle (\theta_A, \theta_B), \theta^{(1)}, \theta^{(2)}, \dots \rangle$. As explained earlier, cheating is feasible only when $\theta_t < \theta_B$, and other things being equal, the incentive to cheat is greater when $\tau_t = A$ than when $\tau_t = B$. So, consider mechanic B in the contingency that $\theta_t < \theta_B$ and $\tau_t = A$ in the initial phase. If he follows the supposed strategy and the initial phase is maintained, the discounted sum of his expected payoff stream is

$$\frac{\delta}{1-\delta}\left((1-p)\int_{\theta_A}^{\infty} \theta dF + \frac{1}{2}\int_{0}^{\theta_A} \theta dF\right) \tag{A1}$$

because he gets zero now and, in each future period t', will provide the repair service with probability $(1-p)$ if $\theta_{t'} \geq \theta_A$ and with probability one-half if $\theta_{t'} < \theta_A$ (i.e., when $\tau_{t'} = B$). On the other hand, if he cheats in this period, he gets θ_t $(< \theta_B)$ with probability $(1-q)$ now, and the first backup phase prevails from the next period onward. This generates a discounted sum of

$$(1-q)\theta_t + \frac{\delta}{2(1-\delta)}\int_0^{\theta^{(1)}} \theta dF. \tag{A2}$$

So, mechanic B has no incentive to cheat if and only if (A1) is at least as large as (A2) for all $\theta_t < \theta_B$. Since (A2) is increasing in θ_t, this is equivalent to

$$\frac{\delta}{1-\delta}\left((1-p)\int_{\theta_A}^{\infty}\theta dF + \frac{1}{2}\int_0^{\theta_A}\theta dF - \frac{1}{2}\int_0^{\theta^{(1)}}\theta dF\right) - (1-q)\theta_B \geq 0. \tag{A3}$$

Next, we find mechanic A's optimality condition. Consider mechanic A, who has learned $\theta_t < \theta_A$ and $\tau_t = B$. For $\theta_t < \theta_B$, the calculation is analogous to that for mechanic B above, from which we find that mechanic A has no incentive to cheat for all $\theta_t < \theta_B$ if and only if

$$\frac{\delta}{1-\delta}\left(p\int_{\theta_A}^{\infty}\theta dF + \frac{1}{2}\int_0^{\theta_A}\theta dF - \frac{1}{2}\int_0^{\theta^{(1)}}\theta dF\right) - q\,\theta_B \geq 0. \tag{A4}$$

For $\theta_t \in [\theta_B, \theta_A)$, however, short-term gain from cheating is greater because he succeeds for sure in this case, while if $\theta_t < \theta_B$, he succeeds only with probability q. It is now a routine calculation to verify that mechanic A has no incentive to cheat for $\theta_t \in [\theta_B, \theta_A)$ if and only if

$$\frac{\delta}{1-\delta}\left(p\int_{\theta_A}^{\infty}\theta dF + \frac{1}{2}\int_0^{\theta_A}\theta dF - \frac{1}{2}\int_0^{\theta^{(1)}}\theta dF\right) - \theta_A \geq 0. \tag{A5}$$

It is straightforward that (A5) implies (A4) because $\theta_A \geq \theta_B$.

So far I have characterized the optimality condition of the agents with (A3) and (A5). However, the special case of $\theta_A = \theta_B = \theta^{(0)}$ is yet to be investigated because, there being no values of θ_t to apply, the inequality (A5) drops out as an optimality condition. In this case, by symmetry, mechanic A's optimality condition coincides with (A3), where $(1-p)$ and $(1-q)$ are replaced by p and q, respectively:

$$\frac{\delta}{1-\delta}\left(p\int_{\theta_A}^{\infty}\theta dF + \frac{1}{2}\int_0^{\theta_A}\theta dF - \frac{1}{2}\int_0^{\theta^{(1)}}\theta dF\right) - q\theta_B \geq 0. \tag{A6}$$

Adding (A3) and (A6) side by side and taking a half of both sides (remember $\theta_A = \theta_B$), we get

$$\frac{\delta}{1-\delta}\left(\frac{1}{2}\int_{\theta_A}^{\infty}\theta dF + \frac{1}{2}\int_0^{\theta_A}\theta dF - \frac{1}{2}\int_0^{\theta^{(1)}}\theta dF\right) - \frac{1}{2}\theta_A \geq 0,$$

which coincides with (A3) and (A6) for the case $p = q = 1/2$. This means that if the agents' optimality is satisfied for a tuple $(\theta_A, \theta_B, p, q)$, then so it does for the tuple $(\theta_A, \tilde{\theta}_B = \theta_A, 1/2, 1/2)$. *Q.E.D.*

Proof of Lemma 4. Below I prove Lemma 4 for the case that the customer always consults both mechanics when $\theta_t = \theta'$. Analogous arguments work for other cases, e.g., when she consults randomly or always consults one mechanic.

Suppose, as in the lemma, that truthful revelation occurs for θ'. Then it must be the case that at least one mechanic, say A, reports truthfully for θ': otherwise, the customer receives obscure messages (in the sense that they may have been sent in either contingencies, $\tau_t = A$ and $\tau_t = B$) from both agents with a positive probability, in which case she cannot hire the right mechanic with certainty, contradicting truthful revelation.

For $\theta_t = \theta'$, let m^A and m^B denote the equilibrium messages that mechanic A sends when $\tau_t = A$ and B, respectively; let n^A, n^B, and n^C denote the ones that mechanic B may send only when $\tau_t = A$, only when $\tau_t = B$, and in either contingency, respectively. Because the right mechanic is hired for sure for θ', the customer's response to the received message pair must satisfy:

(i) hire mechanic $i (= A, B)$ when (m^i, n^i) or (m^i, n^C) is received.

For the remaining two possible message pairs (to be encountered off-equilibrium), she may randomize:

(ii) hire mechanic A with probability r when (m^A, n^B) is received;
(iii) hire mechanic A with probability r' when (m^B, n^A) is received.

Suppose $\theta_t = \theta'$ is realized on the equilibrium path. Let V^A denote the discounted sum of equilibrium payoffs for mechanic A from the next period and onward (valued at the next period). Then the corresponding value for mechanic B is

$V^B = \frac{1}{1-\delta}E(\theta) - V^A$ because the service must be provided by one of the two mechanics in each period. (The equilibrium being stationary, V^A and V^B are indeed the (*ex ante*) discounted sum of equilibrium payoffs.)

If $\tau_t = B$, the expected short-term gain for mechanic A from cheating is at least $r\theta'$. (It is higher if mechanic B sometimes sends n^C.) For honest reporting to be optimal for him in this case, the following is necessary (but generally not sufficient):

$$\delta V^A \geq r\theta'. \tag{A7}$$

If $\tau_t = A$, an analogous argument for mechanic B establishes

$$\frac{\delta}{1-\delta}E(\theta) - \delta V^A \geq (1-r)\theta'. \tag{A8}$$

Adding (A7) and (A8) side by side, I now prove

$$\frac{\delta}{1-\delta}E(\theta) \geq \theta'.$$

Q.E.D.

Proof of Theorem 3. Consider a rivalry sequence $S_r = \langle(\theta^{(0)}, \theta^{(0)}), \theta^{(1)}, \dots\rangle$ as specified in Theorem 3. It is clear that the truncation $S_r^{(1)} = \langle\theta^{(1)}, \theta^{(2)}, \dots\rangle$ constitutes a PAE described in Theorem 1. Hence, it only remains to show the optimality in the initial phase of S_r.

By definition of $\bar{\theta}_r$, each agent's behavior in the initial phase is a best response in S_r if and only if $\theta^{(0)} \leq \bar{\theta}_r(\theta^{(1)})$. As for the customer, she has no incentive to deviate in the hiring decision due to (P4) as in Section 3. With regard to the consultation behavior, it is straightforward to show that she has no incentive to deviate in the initial phase if and only if $\theta^{(0)} \geq \theta^{(1)}$: if $\theta^{(0)} < \theta^{(1)}$, she can maneuver a transition to phase 1 to enjoy more reliable reports, specifically by refusing to provide the report to one mechanic and consulting only him in the next period (see (P1a)). It is straightforward to verify that the customer's behavior in an RAE $S_r = \langle(\theta^{(0)}, \theta^{(0)}), \theta^{(1)}, \dots\rangle$ is indeed optimal if the trust level does not increase, i.e., $\theta^{(k)} \geq \theta^{(k+1)}$ for all $k \geq 0$ as (10) requires. (The explanatory details are analogous to those for Theorem 1 and hence are omitted.) Since $\bar{\theta}_r(\cdot)$ is a decreasing function, the set of effective trust levels sustainable in an RAE is $[0, \bar{\theta}_r(0)]$, where $\bar{\theta}_r(0) = [\delta/(1-\delta)]E(\theta)$ from (9). The RAE with the maximum trust level is clearly $\langle(\bar{\theta}_r(0), \bar{\theta}_r(0)), 0, 0, \dots\rangle$. *Q.E.D.*

Proof of Theorem 6. To find the maximum honesty level sustainable by an n-RAE let $\langle 0, 0, \dots\rangle$ be the sequence of backup phases. Then, optimality in noninitial phases is automatic, and I focus on the initial phase. Consider a panel member, say mechanic A, of an n-RAE with trust level $\theta^{(0)}$. In the case that $\theta_t < \theta^{(0)}$ and $\tau_t \neq A$, for effective cheating he needs to form a collusion consisting of at least $n/2$ members. Since a larger collusion reduces his share of the proceeds from cheating, the most efficient collusion consists of $(n+1)/2$ members if n is odd. If n is even, we need to compare two possibilities: a collusion of $n/2$ members has a $1/2$ chance of success (because the customer will evenly randomize the right mechanic and the mechanic recommended by the collusion), whereas a collusion of $(n/2)+1$ members is assured of success but each member's share is smaller.[22] I examine even-numbered RAE's first and then verify that odd-numbered RAE's perform worse.

Consider an n-RAE where n is an even number. If $n/2$ agents form a collusion, the expected gain from collusive cheating is θ_t/n for each member because they succeed with probability $1/2$, in which case they split the proceeds evenly. If $(n/2)+1$ agents form a collusion, the expected gain from collusive cheating is $2\theta_t/(n+2)$. Because $n \geq 2$, the latter is bigger than the former (the same when $n = 2$).[23] Since the future expected payoff of a cheater is zero in any case, if it is not profitable to form a collusion of $(n/2)+1$ members, neither is it profitable to form any other collusion. I formulate this condition below.

From the above discussion, I calculate that the discounted sum of expected payoffs for a collusion member is $2\theta_t/(n+2)$. The condition that this is lower than that when the initial phase is maintained for all $\theta_t \leq \theta^{(0)}$ is

$$\frac{\delta}{1-\delta}\left(\frac{1}{n}\int_{\theta^{(0)}}^{\infty}\theta dF + \frac{1}{N}\int_{0}^{\theta^{(0)}}\theta dF\right) - \frac{2}{n+2}\theta^{(0)} \geq 0. \tag{A9}$$

Define $\bar{\theta}_n^N(0)$ to be the value of $\theta^{(0)}$ at which (A9) is satisfied tightly (the argument "0" in $\bar{\theta}_n^N(0)$ signifies the backup

[22] If, for example, mechanics A and B are recommended by two and three panel members, respectively, one may argue that the customer should infer that mechanic B is the collusive one because a minority collusion does not make sense. But such an inference would backfire by rendering two-member collusions indeed effective and thereby enhancing the incentive to cheat. The majority rule that I adopted for the hiring decision in case of disagreement is the one that minimizes the incentive to cheat.

[23] When $n = 2$, a collusion of both panel members is absurd. But mathematically it is equally as profitable as the unilateral cheating attempt by either member, so the analysis is unaffected.

trust level): $\bar{\theta}_n^N(0)$ is the highest $\theta^{(0)}$ subject to (A9) because the left-hand side of (A9) is decreasing in $\theta^{(0)}$. Since the optimality of the customer is warranted by nonincreasing trust levels as mentioned earlier, $\bar{\theta}_n^N(0)$ is the maximum trust level sustainable by an n-RAE. The value $\bar{\theta}_r(0)$ in Section 5 is the special case that $N = n = 2$.

Lemma A1. Suppose $N \geq 2$ and n is an even number between 2 and N.
(i) If $N > 2$, then $\bar{\theta}_n^N(0) < \bar{\theta}_r(0)$.
(ii) If $N < N'$, then $\bar{\theta}_n^N(0) > \bar{\theta}_n^{N'}(0)$.

Proof. Part (ii) is immediate from (A9): $\theta^{(0)} = \bar{\theta}_n^N(0)$ violates (A9) for N' and, therefore, part (ii) follows.

Part (i): Since $\bar{\theta}_r(0)$ solves (A9) tightly when $N = n = 2$, we have

$$\frac{\delta}{1-\delta}\left(\frac{2}{n+2}\int_{\bar{\theta}_r(0)}^{\infty}\theta dF + \frac{2}{n+2}\int_0^{\bar{\theta}_r(0)}\theta dF\right) = \frac{2}{n+2}\bar{\theta}_r(0).$$

Since $1/N \leq 1/n \leq 2/(n+2)$ and at most one inequality holds tightly, it follows that

$$\frac{\delta}{1-\delta}\left(\frac{1}{n}\int_{\bar{\theta}_r(0)}^{\infty}\theta dF + \frac{1}{N}\int_0^{\bar{\theta}_r(0)}\theta dF\right) < \frac{2}{n+2}\bar{\theta}_r(0),$$

which violates (A9). Therefore, part (i) follows. *Q.E.D.*

Next, consider odd-numbered n-RAE's. As I said earlier, the most effective collusion size for this case is $(n+1)/2$. This collusion is of the same size as the most effective collusion for an $(n-1)$-RAE discussed above, therefore the expected gain from collusive deviation is the same. But the reward from being faithful is higher in an $(n-1)$-RAE because the customer randomizes among fewer agents for the distrusted range. Hence, a panel member has less incentive to (collusively) deviate in a size $(n-1)$ panel than in a size n panel and, consequently, a higher trust level is sustained by an $(n-1)$-RAE. So,

$$\bar{\theta}_{n-1}^N(0) > \bar{\theta}_n^N(0) \text{ if } n \text{ is odd}.$$

Combining with Lemma A1, this proves the first claim (on the maximum sustainable trust level) of Theorem 6.

Now I prove the claims on recursively credible n-RAE's. I have already established that $\bar{\theta}(\cdot)$ and $\bar{\theta}_r(\cdot) = \bar{\theta}_2^N$ are strictly decreasing functions with a common fixed point θ^*. As for $\bar{\theta}_n^N$ for $n > 3$, note that $\bar{\theta}_n^N(\theta^{(k+1)})$ is the value of $\theta^{(k)}$ that tightly solves

$$\frac{\delta}{1-\delta}\left(\frac{1}{n}\int_{\theta^{(k)}}^{\infty}\theta dF + \frac{1}{N}\int_0^{\theta^{(k)}}\theta dF - \frac{1}{N}\int_0^{\theta^{(k+1)}}\theta dF\right) - \frac{2}{n+2}\theta^{(k)} \geq 0 \tag{A10}$$

when n is even. (When n is odd, the coefficient of the last term is $2/(n+1)$. I present the argument for even n. The same argument works for odd n.) Hence, $\bar{\theta}_n^N$ is a strictly decreasing function. Fix $\theta^{(k+1)} = \theta^*$. If $n = 2$, the value of the left-hand side of (A10) is zero when $\theta^{(k)} = \theta^*$ because θ^* is the fixed point of $\bar{\theta}_2^N$. Compare the left-hand side of (A10) when $n \geq 3$ with the case $n = 2$ for $\theta^{(k)} = \theta^*$:

$$\frac{\delta}{1-\delta}\left(\frac{1}{n}\int_{\theta^*}^{\infty}\theta dF\right) - \frac{2}{n+2}\theta^* < \frac{\delta}{1-\delta}\left(\frac{1}{2}\int_{\theta^*}^{\infty}\theta dF\right) - \frac{2}{2+2}\theta^* = 0.$$

(The inequality follows because $[\delta/(1-\delta)][(1/n)\int_{\theta^*}^{\infty}\theta dF]$ is a larger fraction of $[\delta/(1-\delta)][(1/2)\int_{\theta^*}^{\infty}\theta dF]$ than $[2/(n+2)]\theta^*$ is of $[2/(2+2)]\theta^*$.) This means that (A10) is violated for $n \geq 3$ at $\theta^{(k)}$ given $\theta^{(k+1)} = \theta^*$, which in turn implies that the fixed point of $\bar{\theta}_n^N$ is strictly lower than θ^*.

Given this observation and the fact that the trust level of subsequent phases cannot be lower not to be overriden, I deduce as before that the first two trust levels of a round-k credible n-RAE must be the same number between zero and θ^*, for all $n = 1, \ldots, N$. Then, an argument analogous to that used in the proof of Theorem 2 establishes that $S^* = \langle\theta^*, \theta^*, \ldots\rangle$ is a recursively credible 1-RAE. Similarly, $S_r^* = \langle(\theta^*, \theta^*), \theta^*, \ldots\rangle$ is a recursively credible 2-RAE. In fact, 1- and 2-RAE's consisting of phases with one- or two-member panels with trust level θ^* are all recursively credible.

If an n-RAE has a constant trust level for $n \geq 3$, it is strictly lower than θ^* because $\bar{\theta}_n^N(\theta^*) < \theta^*$ if $n \geq 3$, as shown above, and hence it would be overriden by S^*. Therefore, for $n \geq 3$, a recursively credible n-RAE does not exist. This also proves that any recursively credible 1- or 2-RAE cannot have a panel of three or more members in any phase. *Q.E.D.*

References

ALLEN, F. "Reputation and Product Quality." *RAND Journal of Economics*, Vol. 15 (1984), pp. 311–327.

AUSTEN-SMITH, D. "Information Transmission in Debate." *American Journal of Political Science*, Vol. 34 (1990), pp. 124–152.

BENABOU, R. AND LAROQUE, G. "Using Privileged Information to Manipulate Markets: Insiders, Gurus and Credibility." *Quarterly Journal of Economics*, Vol. 107 (1992), pp. 921–958.

BERNHEIM, B.D., PELEG, B., AND WHINSTON, M.D. "Coalition-Proof Nash Equilibria: Concepts." *Journal of Economic Theory,* Vol. 42 (1987), pp. 1–21.

BLUME, A. "Equilibrium Refinements in Sender-Receiver Games." *Journal of Economic Theory,* Vol. 64 (1994), pp. 66–77.

CROXSON, B., PROPPER, C., AND PERKINS, A. "Do Doctors Respond to Financial Incentives? UK Family Doctors and the GP Fundholder Scheme." *Journal of Public Economics,* Vol. 79 (2001), pp. 375–398.

FERREIRA, J.L. "A Communication-Proof Equilibrium Concept." *Journal of Economic Theory*, Vol. 68 (1996), pp. 249–257.

GARICANO, L. AND SANTOS, T. "Referrals." *American Economic Review*, Vol. 94 (2004), pp. 499–525.

GILLIGAN, T.W. AND KREHBIEL, K. "Asymmetric Information and Legislative Rules with a Heterogeneous Committee." *American Journal of Political Science*, Vol. 33 (1989), pp. 459–490.

HERMALIN, B. "The Boy Who Cried Wolf: Honesty in Repeated Games of Asymmetric Information." Mimeo, University of California at Berkeley, 1998.

KIM, J.-Y. "Cheap Talk and Reputation in Repeated Pretrial Negotiation." *RAND Journal of Economics*, Vol. 27 (1996), pp. 787–802.

KLEIN, B. AND LEFFLER,K.B. "The Role of Market Forces in Assuring Contractual Performances." *Journal of Political Economy*, Vol. 89 (1981), pp. 615–641.

KREPS, D.M. AND WILSON, R. "Reputation and Imperfect Information." *Journal of Economic Theory*, Vol. 27 (1982a), pp. 253–279.

——— AND ———. "Sequential Equilibria." *Econometrica*, Vol. 50 (1982b), pp. 863–894.

KRISHNA, V. AND MORGAN, J. "A Model of Expertise." *Quarterly Journal of Economics*, Vol. 116 (2001), pp. 747–775.

LIPMAN, B.L. AND SEPPI, D.J. "Robust Inference in Communication Games with Partial Provability." *Journal of Economic Theory*, Vol. 66 (1995), pp. 370–405.

MALCOMSON, J.M. "Health Service Gatekeepers." *RAND Journal of Economics*, Vol. 35 (2004), pp. 401–421.

MILGROM, P. AND ROBERTS, J. "Predation, Reputation, and Entry Deterrence." *Journal of Economic Theory,* Vol. 27 (1982), pp. 280–312.

MORRIS, S. "Political Correctness." *Journal of Political Economy*, Vol. 109 (2001), pp. 231–265.

OTTAVIANI, M. AND SORENSEN, P. "Information Aggregation in Debate: Who Should Speak First?" *Journal of Public Economics,* Vol. 81 (2001), pp. 393–421.

PARK, I.-U. "Generic Finiteness of Equilibrium Outcome Distributions for Sender-Receiver Cheap-Talk Games." *Journal of Economic Theory,* Vol. 76 (1997), pp. 431–448.

PESENDORFER, W. AND WOLINSKY, A. "Second Opinions and Price Competition: Inefficiency in the Market for Expert Advice." *Review of Economic Studies*, Vol. 70 (2003), pp. 417–439.

PITCHIK, C. AND SCHOTTER, A. "Honesty in a Model of Strategic Information Transmission." *American Economic Review,* Vol. 77 (1987), pp. 1032–1036.

ROGERSON, W.P. "Reputation and Product Quality." *Bell Journal of Economics*, Vol. 14 (1983), pp. 508–516.

SATTERTHWAITE, M.A. "Consumer Information, Equilibrium Industry Price, and the Number of Sellers." *Bell Journal of Economics*, Vol. 10 (1979), pp. 483–502.

SHAPIRO, C. "Premiums for High Quality Products as Returns to Reputations." *Quarterly Journal of Economics*, Vol. 98 (1983), pp. 659–679.

SHIN, H.S. "The Burden of Proof in a Game of Persuasion." *Journal of Economic Theory,* Vol. 64 (1994), pp. 253–264.

SOBEL, J. "A Theory of Credibility." *Review of Economic Studies*, Vol. 52 (1985), pp. 557–573.

TAYLOR, C.R. "The Economics of Breakdowns, Checkups, and Cures." *Journal of Political Economy*, Vol. 103 (1995), pp. 53–74.

WOLFRAM, S. *The Mathematica Book,* 5th ed. Wolfram Media, Inc., 2003.

WOLINSKY, A. "Competition in a Market for Informed Experts' Services." *RAND Journal of Economics*, Vol. 24 (1993), pp. 380–398.

———. "Eliciting Information from Multiple Experts." *Games and Economic Behavior*, Vol. 41 (2002), pp. 141–160.