Cheap Talk Reputation and Coordination of Differentiated Experts

by

In-Uck Park*

Department of Economics
University of Bristol, U.K.

Department of Economics
University of Pittsburgh, U.S.A.

July, 1999

Abstract

This paper examines the effectiveness of cheap talk advice in recurrent relationships between a customer, and multiple experts who provide professional services with differentiated specialties. Specifically, the sustainable honesty level is characterized in relation to the degree of rivalry among the experts. The three 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 a certain 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 "trusting" them all the time.

JEL Classification Numbers: D82
Keywords: cheap talk, reputation, experts

---

* I am grateful to Andrew McLennan, Jack Ochs, Tuvana Pastine and the participants at the 1999 N. American Winter Meeting of the Econometric Society for their helpful comments. `I.Park@bristol.ac.uk`
Mailing address: 8 Woodland Road, Department of Economics, Bristol BS8 1TN, U.K.

1. Introduction

Customers often have to rely on experts' advice to identify their specific service needs and/or to choose the right service provider. Examples include medical services, repair services, and various consulting services such as those concerning financial investments. Existing studies[1] show that an expert in a recurrent relationship faces a tradeoff between temporary gains from opportunistic behavior and future losses from a damaged reputation. When many experts, each with their own specialty, compete for a customer, there arises an additional consideration: coordination among experts in matching the needed services to the right specialists. While this coordination increases the customers' surplus (and hence social welfare), the experts have no intrinsic interest in it because they are engaged in a zero sum game amongst themselves. Therefore, it is up to the customer to discipline them to improve the coordination by extracting trustworthy advice or referrals. Indeed, many such relationships in practice appear to be based on high levels of mutual trust: a customer patronizes an expert out of trust, and an expert behaves faithfully trusting that the customer will return for future patronage.

This paper studies such trust on cheap talk[2] advice in recurrent relationships between a customer and multiple experts with differentiated specialties. We find this issue particularly interesting because, although cheap talk consultation is prevalent in many such environments, it has an intrinsic weakness as a means of information transmission due to the costless nature. We feel, therefore, that trust can be vital for effective cheap talk communication.

Specifically, we characterize the sustainable honesty level of the experts' advice in relation to the degree of rivalry among the experts. There are three main findings. 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 a certain 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 "trusting" them all the time.

We illustrate the main intuition for these results in a context of car repair services. Consider a car owner in a town with two mechanics. The problems with the car are classified into two types, say A and B, and each mechanic provides high quality services for problems of one type (his specialty) and low quality services for the other type. The quality of the service is known to the customer only after the purchase. Each time a problem occurs, the car owner wishes to hire the right mechanic, but she does not know the type of the problem. So, she consults the mechanics for free cheap talk advice. Both mechan-

[1] See, for example, Sobel (1985), Kim (1996) and Morris (1998).

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

ics, however, have incentives to claim that the problem is their specialty for short-term gain. The question is how and to what extent the customer can restore honest advice by provoking reputational motives, which is an inherent disciplinary means in repeated relationships. To focus on this issue, other interesting aspects of the considered environment, such as search costs, price competition and legal liabilities, are set aside.

Full honesty is generally not obtained, because the mechanic would try to cash in his reputation by misleading the customer if the profitability of the current problem (which is a random draw) is sufficiently high. It turns out that for the purpose of investigating the sustainable honesty levels, it suffices to focus on two categories of equilibria. The first category depicts situations in which the customer patronizes one of the mechanics as her "primary" agent as long as he lives up to her expectation; once he fails it, the customer switches to make the other mechanic her new primary agent (referred to as the "backup" agent). Since full honesty is not sustainable, there is some level of dishonesty that the customer tolerates while continuing to patronize the (original) primary agent,[3] which generates "bonuses" or extra profits for him. Such a generous treatment of the primary agent by the customer enhances her value as a future customer, thus disciplining him because of the high cost of losing a valued customer. On the other hand, having the full trust of the customer, the primary agent has a strong temptation to cheat because success is guaranteed.

The second category depicts situations in which the customer holds the two mechanics in check by consulting them both and randomly hiring one or the other if they differ in their advice. As before, the customer tolerates a certain level of dishonesty from both mechanics; if she detects an incidence of dishonesty beyond the tolerated level, she punishes the dishonest mechanic by adopting the other mechanic as her primary agent (the backup agent) in the manner explained above. Because the mechanics "share" the customer in this category of equilibria, her value as a future customer to each mechanic is lower than it is to the primary agent in the first category. At the same time, the expected opportunistic gain from cheating is also lower because the success rate is lower (due to randomization). The latter effect is dominant when the honesty level expected from the backup agent is low. In addition, a low backup honesty level by itself pushes up the initial honesty level by rendering the punishment severe (because the backup agent will be treated more generously at the expense of the cheater). Hence, the maximum honesty level sustainable in this category is higher than that in the first category. In fact, that level is shown to be the upper bound of the honesty level in *any* equilibrium.

But such maximum honesty levels are supported by an extreme punishment threat that a mechanic will never be hired (let alone consulted) again if he ever cheats. This threat does not sound very credible because once punishment starts, rather than blindly hiring the non-cheater all the time to keep her threat, the customer would be interested in, for example, a new deal with the cheater (and even with the non-cheater, using her bargaining

---

[3] Casual observations seem to support this phenomenon.

power) in which he would behave more honestly in return for being adopted again as the primary agent. However, internal consistency requires that to validly overturn the original equilibrium, such deviations in punishment phase themselves need be robust to the same credibility check. The equilibria that pass such a credibility criterion on the punishment phase, formalized as "recursive credibility," implement the same (unique) honesty level in the two categories explained above (see below for intuitions).

We extend the analysis to cases of more than two experts with their own specialties, where the customer may consult any subset of experts, called a "panel", and a majority of them may cheat collusively, i.e., coordinate to mislead the customer and split the proceeds.[4] Again, a tolerated level of dishonesty generates bonuses to the panel, and a panel member is forfeited his membership if he cheats beyond the tolerated level. We show that the sustainable level of honesty deteriorates as there are more experts. This is because the customer is worth less to each expert because he provides the service less frequently due to finer specialization: then there are greater incentives to cheat because there is less to lose. Nonetheless, the "recursively credible" equilibria still implement the same honesty level regardless of the number of experts. Furthermore, this level can be achieved by trusting a panel of only one or two experts all the time.[5]

The intuition for the last two results are as follows. Since every period has the same continuation game in an infinitely repeated setting, internal consistency basically means that what can be arranged today can be arranged tomorrow, and vice versa. So, the initial and the backup honesty levels are identical in "recursively credible" equilibria. In this case, the cheating incentives of a collusion member are determined by the balance between his share of the expected short-term gain from cheating and the foregone future share of bonuses as a panel member.[6] These are determined by collusion size and success rate, and by panel size, respectively. In particular, the total number of experts does not affect the incentives and so, not the honesty level, either.

For a one- or two-member panel an effective collusion is the whole panel, while for a larger panel an efficient collusion is just over a half of the panel. The cheating incentives are greater for the latter because the proceeds are shared by relatively fewer experts than the bonuses are. As a result, a panel of one or two experts supports a higher honesty level.

The model is pertinent to various other situations. For example, financial consultants may have expertise in different areas (e.g, in stock investments or in pensions and insurance) and may have varying motives each time (e.g, off-loading surplus stock or meeting quarterly targets), and the clients may be ignorant as to which investments are suitable for their current situations. This paper provides some guidelines for efficient strategies of

---

[4] Mutual referrals within a small circle of professional service providers may be collusive behavior.

[5] Satterthwaite (1979) shows that an increased number of sellers (experts) in a monopolistically competitive market of a "reputation good" may cause the price to rise. Central to his result is the search cost that increases as there are more sellers, causing individual seller's demand less price elastic.

[6] If the two honesty levels differ, his "regular" service profit from honest advice also needs to be taken into account because it is affected by cheating. This is sensitive to the total number of experts.

the customer in such situations. In fact, it provides an explanation from an incentives perspective (rather than search costs) for the prevalent practice of patronizing a small number of experts in such environments. In addition, this paper contains some implications for efficient organization of private or public enterprises providing professional services. The health care market appears as a particularly interesting case because a primary care system and a self-referral system co-exist.

This paper complements the existing cheap talk literature (reviewed below) in two respects. Firstly, this paper deals with situations in which multiple experts cooperate in promoting social welfare by giving more reliable advice; in most other studies with multiple experts, no room for such cooperation exists and each expert tries to influence the decision maker at the expense of the other. Secondly, since the identities of the experts (the right or wrong mechanic) are reset independently each time, the standard reputational argument (based on (mis-)learning the fixed, true identity over time via observed behavior) does not apply. Instead, reputation in this model is sustained by self-disciplinary behavior which credibly signals the committment to reputational behavior in the future.[7]

Related literature and organization of paper

This paper contributes to the literature on cheap talk reputation. Sobel (1985) shows that an "enemy" (an informed agent with completely opposing interests to the decision maker) may build reputation by mimicking the honest behavior of a "friend" (with identical interests 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 political context, Morris (1998) shows that even a friend may have incentive to lie if the signal is noisy and he is sufficiently concerned about his reputation. In these studies, 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 (like this paper), Kim (1996) shows that infinitely repeated pretrial negotiation can enhance the credibility of cheap talk and improve efficiency. This paper complements Sobel (1985), and extends Kim (1996). Ottaviani and Sorensen (1999a,b) also study cheap talk reputation but the experts in their model are motivated by exogenous reputational payoff that is increasing in estimated ability level, *a la* the career concerns literature.

The role of cheap talk advising has been explored in the provision of credence services, too. In particular, Pitchik and Schotter (1987) examine the honesty level of cheap talk advice in a one-shot game of credence good provision.[8]

Multiple experts with conflicting interests have been investigated in static (i.e, one-

---

[7] A similar basic idea has been explored by Klein and Leffler (1981) and Shapiro (1983) in the context of a repeat-purchase goods market.

[8] A credence good, due to Darby and Karni (1973), is one for which quality may never be known to the purchaser. (The service in the current paper is an experience, rather than credence, good because the quality is revealed ex-post.) See also, for example, Wolinsky (1993) and Taylor (1995) for credence goods markets.

shot) settings. Gilligan and Krehbiel (1989) and Austen-Smith (1990) examine the effects of multiple representatives with their own agendas in varying legislative procedures. Shin (1994) examines the decision rule in an arbitration process when the information partitions of the informed parties are uncertain although their reports are verifiable. Lipman and Seppi (1995) show that limited provability may extract full information when speakers with conflicting preferences talk sequentially. In a model of the provision of credence services where the diagnostic effort is unobservable, Pesendorfer and Wolinsky (1998) show that price competition of identical experts results in inefficiency. Krishna and Morgan (1999) study efficient information extraction from two experts in relation to the directions of their biases relative to the decision maker. Dewatripont and Tirole (1999) investigate advocacy in the context of a moral hazard model.

The rest of the paper is organized as follows. Section 2 describes the model of two experts. Sections 3 and 4 formally examine, respectively, the two categories of equilibria discussed above. Section 5 extends the analysis to cases with more than two experts. Appendix contains missed proofs.

2. Model and Preliminaries

There are an infinite sequence of periods indexed by $t = 1, 2, \cdots$, and three long-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 which has to be repaired by one of the two mechanics. This problem is characterized by two independent random variables: its *type* τ_t is either A or B with even probabilities and its *importance* θ_t is realized according to a probability distribution function $F(\theta)$ and density $f(\theta)$ supported on $\Re_+$. The type represents the nature of the problem and mechanic A (B) is better at repairing problems of type A (B). The importance parameter θ_t measures the seriousness of the problem of that period and determines the profit level of the mechanic who performs the repair service. We assume that the expected value of θ, $E(\theta) = \int_0^\infty \theta dF$, is finite.

The stage game proceeds as follows. When a problem occurs in period t, the customer knows the value of its importance θ_t, but not its type τ_t. She *consults* either mechanic A or B (possibly, both) for a diagnose. Either mechanic correctly identifies the values of τ_t and θ_t when consulted, and sends a cheap talk message regarding the type of the problem. Based on the messages received, the customer updates her belief on the problem's type and hires a mechanic for repair service.

We set aside the issue of search cost by assuming that there is no cost for either mechanic to identify τ_t and report about it and, therefore, the consultation is free of charge.[9] However, consultation activity is assumed to be private between each mechanic and the customer so that the mechanics can not base their reports on whether the other mechanic had been consulted. This is basically to avoid possible analytic complications due to the first or last mover's advantage between the two mechanics, which does not appear essential in many circumstances.

[9] Small search costs do not change the qualitative results.

We now specify the payoff structure. Because our main concern is the effectiveness of cheap talk in the presence of multiple informed agents (the mechanics) vying for a lasting relationship with a principal (the customer), we do not allow the mechanics to compete in ways other than cheap talk (such as in price). In particular, the payoffs to the players in period t are completely determined by τ_t, θ_t and who performed the repair service in period t. A mechanic gets a payoff θ_t if he performed the repair service and gets 0 otherwise. The customer gets a payoff of u if the problem is repaired by the mechanic τ_t, the specialist for the current type; she gets 0 otherwise. These payoffs are summarized in the next table where the payoffs are listed for mechanic A, mechanic B and the customer in that order:

	A	B
$\tau = A$	$\theta, 0, u$	$0, \theta, 0$
$\tau = B$	$\theta, 0, 0$	$0, \theta, u$

We examine the situation that 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 paper remain valid in the case that θ_t is revealed to the customer at the end (rather than at the beginning) of period t. The same is true when the customer's payoffs also depend on θ (for example, θu or u/θ instead of u) as long as she prefers mechanic A (B) when type A (B) is more likely than the prior.

The assumption of even prior on the problem's type τ_t, however, is important in our analysis and discussions. For instance, the rivalry between the two mechanics would not be on the level playing field if the customer is biased to one of the mechanics. It does not seem very realistic to assume that the customer truly believes that every sort of problem arises with exactly the same probability. A more sensible interpretation of the even prior assumption would be that sometimes the customer knows the type of the problem and needs no consultation, and other times the problem is too complicated or new for the customer to self-diagnose. The paper models repeated occurences of the latter sort of problems, for which an unbiased prior seems plausible.

The key element in the considered environment is the informational contents of the cheap talk messages sent by the mechanics, which in equilibrium are determined by Bayesian updating. In a one shot-game, these messages carry no value due to an intrinsic conflict of interests between the customer and the mechanics: mechanic A has every intention to truthfully report type A problems by sending particular messages (with the aim of inducing the customer to hire him), but when the problem is type B he would still send the same messages with the aim of misleading the customer to believe the problem to be of type A and to hire him. Hence, the messages sent by mechanic A carry no informational content, and the customer attaches no meaning to them. The same applies to mechanic B and there is no room for cheap talk communication. 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.[10]

We focus on more interesting equilibria of the repeated framework in which effective cheap talk communication arises by the consideration of reputation. However, fully honest reporting can not be sustained: if the current value of θ_t is sufficiently high the consulted mechanic has incentives to mislead the customer, because the opportunistic gain would overcompensate the discounted sum of future losses in payoff stream from losing the customer's trust. In equilibrium the customer would take these incentives into account and interpret the messages as meaningless. For other values of θ_t, on the other hand, the potential opportunistic gain would not justify future losses and so the consulted mechanic would report honestly by sending a particular message if $\tau_t = A$ and another distinct message if $\tau_t = B$. We 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 the above discussion, it appears most natural for each mechanic to adopt a cut-off strategy in each period t, if consulted: he reports honestly if $\theta_t < \tilde{\theta}_t$ for a certain critical 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*). We say that a mechanic *reports with a trust level* $\tilde{\theta}_t$ if he uses this strategy. We say that a mechanic *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 trust level of each mechanic may vary from period to period in equilibrium, provided that such variation is correctly anticipated by the customer. However, since in each period the players face exactly identical future there seems to be no sensible reason for the trust level to vary between periods, unless a deviation has taken place. In this paper, therefore, we consider "stationary" equilibria in which the players report with the same trust level in periods t and $t' (> t)$ if there has been no deviation in between, that is, in periods $t, t+1, \cdots, t'-1$. At the end of Section 4 we show that this class of equilibria effectively covers all equilibria, in the sense that for any other equilibrium there exists a stationary one with the same consumer's and mechanics' surpluses.

3. Primary Agency Equilibrium

We 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) he 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. If the customer patronizes a trusted agent, she is more vulnerable to cheating, but future punishment is greater because the future business at stake is bigger.

A primary agency equilibrium consists of a sequence of phases, each with a trusted agent. Phase 0, or an *initial phase,* comprises of periods $t = 1, 2, \cdots$, in which the customer is supposed to patronize one of the mechanics as the trusted agent with an initial trust

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

level $\theta^{(0)}$. The trusted agent in the initial phase is called the *primary agent.* With no loss of generality, let mechanic A be the primary agent.

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. Then, phase 1, or a *first backup phase,* starts and prevails in periods $t+1, t+2, \cdots$, in which the customer patronizes the other mechanic (mechanic B) as the trusted agent, called the first backup 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 $t+1$,[11] and behaves accordingly.

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 $t'+1, t'+2, \cdots$, in which the customer patronizes the mechanic other than the first backup mechanic as the new trusted agent, called the second backup mechanic, with a second backup trust level $\theta^{(2)}$. Higher order backup phases, phases $k = 3, 4, \cdots$, are modelled in an analogous manner, with trust levels $\theta^{(k)}$.

We denote the players' behavior in succesive phases described above by an infinite sequence of nonnegative trust levels $S = \langle \theta^{(0)}, \theta^{(1)}, \cdots \rangle$. The behavior of the continuation game at the beginning of phase $k \geq 1$ is denoted by a truncated sequence $S^{(k)} = \langle \theta^{(k)}, \theta^{(k+1)}, \cdots \rangle$ from phase k and onwards, 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* (p.a.e., hereafter) is an infinite sequence $S = \langle \theta^{(0)}, \theta^{(1)}, \cdots \rangle$ such that each player's behavior is a best response to those of other players in S and in each $S^{(k)}$ for $k = 1, 2, \cdots$. We note, however, that the current description of players' behavior is incomplete because it does not specify off the equilibrium paths when the customer deviates. For expositional convenience, we discuss them later as needed.

To characterize p.a.e, we start with the best response condition for the trusted agent's strategy in each phase. Specifically, we check if it would ever be profitable for the trusted agent to cheat in any period t of phase k. Note that cheating is feasible only when $\theta_t < \theta^{(k)}$: otherwise any report is an equilibrium message because he is supposed to babble anyway. So, consider the trusted agent in an arbitrary period t of phase k, who has been consulted and learned the values $\theta_t < \theta^{(k)}$ and τ_t. With no loss of generality, let mechanic A be the trusted agent of phase k.

If he abides by the supposed strategy of reporting with trust level $\theta^{(k)}$ throughout, phase k will prevail in the future and his expected payoff in each future period (period $t+1$ and onwards) is

$$V(\theta^{(k)}) \equiv \int_{\theta^{(k)}}^{\infty} \theta dF + \frac{1}{2} \int_{0}^{\theta^{(k)}} \theta dF \tag{1}$$

On the other hand, if he cheats in period t, phase $(k+1)$ would start and prevail in the future, so that his expected payoff would be $\frac{1}{2} \int_0^{\theta^{(k+1)}} \theta dF$ in each future period. The differ-

[11] See (P2) below.

ence in the discounted sum of these two streams of future payoffs is the future punishment of cheating.

Next, check the current gain from cheating. If $\tau_t = B$, mechanic A would reap a current gain of θ_t by cheating. If $\tau_t = A$, however, he loses by cheating because he lets mechanic B provide the service when he himself should.

Therefore, mechanic A would never have an incentive to cheat if and only if the current gain when $\tau_t = B$ does not exceed the future punishment for all $\theta_t < \theta^{(k)}$. Since the current gain is higher for higher θ_t, this condition is written as

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

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 (LHS, hereafter) 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 non-trusted agent's behavior is trivially optimal because he does not make any strategic moves, we summarize the agents' optimality in Lemma 1 below. We 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 $S = \langle \theta^{(0)}, \theta^{(1)}, \cdots \rangle$ and in each truncated sequence $S^{(k)}$, if and only if*

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

Lemma 2: *The function $\bar{\theta}(\cdot)$ is a strictly decreasing function. Denoting the unique fixed point of $\bar{\theta}(\cdot)$ 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 $V(\theta^{(k)})$ is strictly decreasing in $\theta^{(k)}$ and hence, so is $W(\theta^{(k)}) = \frac{\delta}{1-\delta}V(\theta^{(k)}) - \theta^{(k)}$. 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 deduce $\bar{\theta}(0) > 0$. Finally, note that since the LHS 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 $\bar{\theta}(0)$. QED

We now move on to the optimality of the customer's behavior in $S = \langle \theta^{(0)}, \theta^{(1)}, \cdots \rangle$ and ask if a deviation would be profitable for the customer. The answer to this question, however, hinges on what would happen after such deviations, which we left unspecified up to now. Specification of such off the equilibrium paths that supports the customer's behavior as a best response, is not unique. Below we describe one specification which, considering the equilibrium behavior, we believe is sensible. We retain (3) in this discussion. We note that off the equilibrium behavior we postulate in this and later analyses can be verified in a straightforward way to be compatible with a "consistent assessment" of Kreps and Wilson (1982). The explanation, however, is lengthy and so is omitted.

When a deviation takes place, the players change their beliefs about future course of the game. We 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 $S = \langle \theta^{(0)}, \theta^{(1)}, \cdots \rangle$, we say that a mechanic *assumes* phase k in period t (actively if he is the k-th backup mechanic of S, and passively if not).

The customer may deviate from S either in consultation or in hiring decision. First, we 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 non-trusted agent, mechanic B, gets consulted in period 1, he assumes phase 1 in period 1.

(P2) Suppose that phase $k(= 0, 1, \cdots)$ 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 non-trusted 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 $\langle \hat{\theta}^{(k)}, \theta^{(k+1)}, \theta^{(k+2)}, \cdots \rangle$ actively in period $t+1$ where $\hat{\theta}^{(k)} = \min\{\theta^{(k)}, \theta^{(k+1)}\}$.[12] 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 $t+1$.[13]

The other kind of possible deviations by the customer is that she may not follow the trusted agent's recommendation in her hiring decision. We postulate:

(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 on the prevailing phase.

In light of (P4), the postulates (P1)$\sim$(P3) also cover the cases that deviations in consultation are preceded by deviations in hiring decisions of previous periods. If the customer makes multiple deviations in consultation over time, each agent updates his

---

[12] This is as if he believes to have been given a second chance. We take the minimum here not to give the customer an incentive to forgive him. Due to (7) to be derived later, this amounts to taking $\hat{\theta}^{(k)} = \theta^{(k+1)}$. If we postulate $\hat{\theta}^{(k)} = \theta^{(k)}$ instead, the "recursively credible" equilibrium (to be discussed later) obtains.

[13] 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 hiring decision, which does not change agents' beliefs as postualted in (P4) below.

belief on the prevailing phase according to the relevant postulate at each incidence of deviation. We note that agents may not have synchronized beleifs on off the equilibrium paths because they may diverge in detecting deviations. For example, if both agents are consulted in phase 0, machanic A would believe to be in phase 0 in the next period while mechanic B would assume phase 1.

Nonetheless, in each period each agent believes to be in the initial phase of a sequence that satisfies (3): this is obvious because they believe to be in some phase of the original sequence S except for mechanic A described in (P3), in which case (3) follows because $\hat{\theta}^{(k)} \leq \theta^{(k)} \leq \bar{\theta}(\theta^{(k+1)})$. Therefore, by Lemma 1, each agent's behavior postulated above is a best response given his belief at that point in time. In addition, apart from one exception, each agent's belief is "consistent" in the sense that it is possible that the other agent has the same belief as his (i.e, there exists a path that is consistent with his experience and belief, and would have lead the other agent to have the same belief, too). The exception is mechanic A in a situation described in (P3) when $\hat{\theta}^{(k)} \neq \theta^{(k)}$. His belief in this case is still "strategically consistent," because his strategic incentives are the same whether mechanic B has the same belief as his or the belief described at the end of (P3).

Having specified the beliefs on off the equilibrium paths following the customer's deviations as above, we now examine the optimality of the customer's equilibrium behavior. Since the future is not affected by current hiring decision (see (P4)), following the recommendation of the trusted agent is obviously optimal hiring decision on the equilibrium paths. Off the equilibrium paths, the customer can keep track of the updating of each agent's belief on the prevailing phase accurately, because it is based on the customer's consultation decisions and the agent's deviations which the customer detects at the end of each period. We postulate that the customer follows the recommendation of the agent who reports with a higher trust level in each period t, which clearly is optimal hiring decision.

Next, we examine optimality of the customer's consultation behavior in each phase k. If the customer maintains phase k she will get an expected payoff of $U(\theta^{(k)})$ in each future period, where

$$U(\theta') \;=\; uF(\theta') + \frac{u}{2}(1 - F(\theta')) \tag{6}$$

because she always gets u by hiring the right mechanic in the trusted range, while in the distrusted range she does so only a half of the time.

However, the customer may deviate by consulting only the non-trusted agent, to induce both agents to assume $\langle \theta^{(k+1)}, \theta^{(k+2)}, \cdots \rangle$ according to (P2). If $\theta^{(k)} < \theta^{(k+1)}$ she would actually do this and patronize the trusted mechanic of phase $(k+1)$ forever, attaining a higher expected payoff of $U(\theta^{(k+1)})$ in each future period. Therefore, the following is necessary for the customer not to deviate in any phase of S:

$$\theta^{(k)} \geq \theta^{(k+1)} \qquad \forall k = 0, 1, 2, \cdots \tag{7}$$

To check sufficiency, suppose (7) holds and consider the customer in period t of phase k. If she maintains phase k, she will get an expected payoff $U(\theta^{(k)})$ in every period. If she deviates in period t, according to (P1)~(P4), in any future period each agent would report (if consulted) with a trust level $\theta^{(k')}$ for some $k' \geq k$. By (7), the maximum expected payoff that the customer can derive from such reports is at most $U(\theta^{(k)})$ and, therefore, we conclude that the customer would never deviate. So, we have

Lemma 3: *Given a sequence $S = \langle \theta^{(0)}, \theta^{(1)}, \cdots \rangle$, augmented by off the equilibrium behavior as specified in (P1)~(P4), the customer's behavior is a best response in each phase if (3) and (7) hold, or equivalently, if*

$$\theta^{(k+1)} \leq \theta^{(k)} \leq \bar{\theta}(\theta^{(k+1)}) \qquad \forall k = 0, 1, 2, \cdots \tag{8}$$

Since the initial phase prevails forever, the trust level actually exercised in a p.a.e. is the initial one, $\theta^{(0)}$. Because $\bar{\theta}(\cdot)$ is a decreasing function, the set of trust levels sustainable by p.a.e. is $[0, \bar{\theta}(0)]$. The p.a.e. with the maximum trust level is $\langle \bar{\theta}(0), 0, 0, \cdots \rangle$. The next theorem summarizes the findings so far.

Theorem 4: *A sequence $S = \langle \theta^{(0)}, \theta^{(1)}, \cdots \rangle$, augmented by off the equilibrium behavior as specified in (P1)~(P4), constitutes a p.a.e. if and only if (8) holds. The set of trust levels sustainable in p.a.e. is $[0, \bar{\theta}(0)]$.*

We find, however, that the equilibrium condition (8) 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 $k = 1, 2, \cdots$. In other words, it is supported by the extreme threat to the primary agent that he 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 non-trusted agent may approach the customer and offer a "coalitional deviation" to start another p.a.e. 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 if it is 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 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.

Our notion of credibility is a variant of coalition-proofness of Bernheim, Peleg and Whinston (1987). Their notion is also recursive but they developed it for cases with finite recursion. In our environment the recursion is inherently infinite and the definition is circular. Nonetheless, it allows us to identify the unique p.a.e. that conforms to the definition.

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

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

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

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

  ii) there does not exist a *round-k credible* p.a.e. that overrides $S^{(1)}$.

(c) A p.a.e. S is *recursively credible* if it is round-k credible for all $k = 1, 2, \cdots$.

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

However, due to circularity of the definition, we cannot check the credibility of an individual p.a.e. separately: round-k credibility of a p.a.e. depends upon that of other p.a.e.'s, and vice versa. Instead, we need to find the sets of round-k credible p.a.e.'s, inductively on k, and then take the intersection to obtain the set of recursively credible p.a.e.'s. Rather than going through the full process,[14] we take a shortcut to identify a recursively credible p.a.e. which turns out to be the unique one.

A round-k credible p.a.e. $S = \langle \theta^{(0)}, \theta^{(1)}, \cdots \rangle$ can not have $\theta^{(0)} > \theta^{(1)}$, because if so, S itself overrides $S^{(1)}$, contradicting condition ii) of part (b) above. Together with condition (7) of p.a.e., it follows that $\theta^{(0)} = \theta^{(1)}$. Since this holds for every k and every truncation of a recursively credible p.a.e. is also recursively credible by definition, any recursively credible p.a.e. must have the same trust level, say θ', for all phases.

For such a p.a.e, (8) implies $\theta' \leq \bar{\theta}(\theta')$. Since $\bar{\theta}(\cdot)$ is decreasing with the fixed point θ^*, we further deduce that candidates for recursively credible p.a.e. are constant sequences of a trust level between 0 and θ^*. Among those, $S^* = \langle \theta^*, \theta^*, \cdots \rangle$ overrides others and is most preferred by the customer. Indeed, we have

Theorem 5: *$S^* = \langle \theta^*, \theta^*, \cdots \rangle$ is the unique recursively credible p.a.e.*

Proof: As discussed above, the first two trust levels of a round-k credible p.a.e. must be the same number between 0 and θ^*.

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

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

---

[14] For example, the set RC(1) of round-1 credible p.a.e.'s consists of the ones with identical initial and the first backup trust levels at a particular level, say θ' (see the first two sentences of the next paragraph). From (8), $\theta' \leq \theta^*$ follows. But, if $\theta' < \theta^*$ then $S^* = \langle \theta^*, \theta^*, \cdots \rangle$ is round-1 credible according to Definition 1(a), resulting in a contradictory conclusion that the elements of RC(1) are not round-1 credible because their truncations are overriden by S^*. Hence, RC(1) consists of p.a.e.'s with $\theta^{(0)} = \theta^{(1)} = \theta^*$. Inductively, one can easily show that RC(k) consists of the ones with $\theta^{(k')} = \theta^*$ for $k' = 0, \cdots, k$.

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

4. Rivalry Agency Equilibrium

In a rivalry agency equilibrium, the customer does not rely on a primary agent in the initial phase but she makes hiring decision based on both mechanics' reports. Formally, an initial phase comprises of periods $t = 1, 2, \cdots$, in which mechanics A and B report with initial trust levels θ_A and θ_B, respectively, where we 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, *iii*) if $\theta_B > \theta_t$, she hires the recommended mechanic if the recommendations coincide, but in case they do not coincide she hires mechanics A and B with probabilities q and $1-q$, respectively.[15]

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 $t+1, t+2, \cdots$, in which the customer patronizes mechanic B as the trusted agent (backup agent) who reports with a first backup trust level $\theta^{(1)}$, i.e, in the same manner as in a p.a.e. explained in Section 3. Higher order backup phases are modelled in the same manner, too.

Transition to the first backup phase needs some further explanation, because it may not be synchronized among all three players. 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 $\theta^{(1)}$.

To circumvent the analytical complication due to such possibilities, we adopt the following assumption:

(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.[16]

In the remainder we 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. We also postulate that each mechanic takes the customer's refusal to provide the other mechanic's report as an evidence of cheating by him. Then, the transition to the first backup phase is unambiguously coordinated by all three players. We note, however, that most of the main results in this

---

[15] p and q can be functions of θ_t.

[16] This is weaker than assuming observability of the report because they can be withheld. The reports are still cheap talk messages because the mechanics are not held responsible for their reports, for instance, in a court.

paper can be obtained without assuming (P5)[17] but at a cost of more complicated off the equilibrium behavior.

We denote the players' behavior in successive phases described above by a modified sequence $S_r = \langle (\theta_A, \theta_B), \theta^{(1)}, \theta^{(2)}, \cdots \rangle$ which we refer to as a *rivalry sequence.* A truncated sequence $S_r^{(k)} = \langle \theta^{(k)}, \theta^{(k+1)}, \cdots \rangle$ of S_r from phase $k\,(\geq 1)$ and onwards, constitutes a sequence that we considered for p.a.e. in the previous section. A *rivalry agency equilibrium* (r.a.e., hereafter) is a rivalry sequence $S_r = \langle (\theta_A, \theta_B), \theta^{(1)}, \theta^{(2)}, \cdots \rangle$ such that each player's behavior is a best response to those of other players in S_r and in each $S_r^{(k)}$ for $k = 1, 2, \cdots$. Since the initial phase prevails forever, the effective trust level of an r.a.e. is θ_A, the higher of the two initial trust levels. An r.a.e. is *symmetric* if $\theta_A = \theta_B$ and $p = q = \frac{1}{2}$.

By definition, the backup phase truncation, $S_r^{(1)} = \langle \theta^{(1)}, \theta^{(2)}, \cdots \rangle$, of an r.a.e. S_r, constitutes a p.a.e. described in Theorem 4. Hence, in the remainder we make it a custom that the backup phase truncation of a rivalry sequence is a p.a.e. In particular, we take (3) and (7) for granted for $k = 1, 2, \cdots$. Then, the best response property is automatic in every backup phase. Below, we focus on the initial phase.

First, in the next lemma we derive a result that the effective trust level of any r.a.e. is implementable by a symmetric r.a.e. 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, would have a greater incentive to cheat, and ii) given $\theta_A = \theta_B$, unequal treatment (i.e, $p \neq \frac{1}{2}$ or $q \neq \frac{1}{2}$) would increase the incentive to cheat for the less favorably treated mechanic and consequently, lower the effective trust level. A detailed proof is provided in Appendix A.

Lemma 6: *Suppose that each agent's behavior is a best reponse in a rivalry sequence* $\langle (\theta_A, \theta_B), \theta^{(1)}, \theta^{(2)}, \cdots \rangle$ *for some values of* p *and* q*. Then, so it is in a symmetric rivalry sequence* $\langle (\theta_A, \hat{\theta}_B = \theta_A), \theta^{(1)}, \theta^{(2)}, \cdots \rangle$ *for* $p = q = \frac{1}{2}$.

In light of Lemma 6, we focus on symmetric r.a.e. from now on and denote the common initial trust level by $\theta^{(0)}$. To check the optimality of the agent's behavior in the initial phase, we pick any agent (because they are symmetric), say mechanic A, and ask if it would ever be profitable for him to cheat in any period t of the initial phase. As explained earlier, cheating is feasible only when $\theta_t < \theta_A$, and other things being equal, the incentive to cheat is greater when $\tau_t = B$ than when $\tau_t = A$.

Hence, consider mechanic A in period t of the initial phase, who examined the car and learned the values $\theta_t < \theta_A$ and $\tau_t = B$. Compared with the case that he is the sole trusted agent (which has been analyzed in Section 3), there are two differences: i) the probability of success is only $\frac{1}{2}$ if he cheats, and ii) he gets to provide the service with a probability $\frac{1}{2}$ when $\theta_{t'} \geq \theta^{(0)}$ in each future period t' if he does not cheat. The best response condition,

[17] The only notable difference is that $\bar{\theta}_r(\theta^{(1)})$ to be derived below coincides with $\bar{\theta}(\cdot)$ that has been derived in Section 3 for p.a.e.

therefore, is a variant of the inequality (2) that accommodates these two differences: either agent would never have an incentive to cheat in the initial period if and only if

$$\frac{\delta}{1-\delta}\Big(\frac{1}{2}\int_0^{\infty}\theta dF - \frac{1}{2}\int_0^{\theta^{(1)}}\theta dF\Big) - \frac{1}{2}\theta^{(0)} \;\geq\; 0 \tag{9}$$

Given $\theta^{(1)}$, define $\bar{\theta}_r(\theta^{(1)})$ to be the value of $\theta^{(0)}$ at which (9) is satisfied tightly:

$$\bar{\theta}_r(\theta^{(1)}) \;=\; \frac{\delta}{1-\delta}\Big(E(\theta) - \int_0^{\theta^{(1)}}\theta dF\Big) \tag{10}$$

Then, $\bar{\theta}_r(\cdot)$ is a well-defined, decreasing function with the property that (9) holds if and only if $\theta^{(0)} \leq \bar{\theta}_r(\theta^{(1)})$. The next lemma summarizes the agents' optimality. The first two equalities of (12) follow because (2) and (9) are equivalent when $\theta^{(1)} = \theta^{(0)}$.

Lemma 7: *Given a rivalry sequence $S_r = \langle(\theta^{(0)}, \theta^{(0)}), \theta^{(1)}, \cdots\rangle$, each agent's behavior is a best response in S_r if and only if*

$$0 \;\leq\; \theta^{(0)} \;\leq\; \bar{\theta}_r(\theta^{(1)}) \tag{11}$$

In addition,

$$\bar{\theta}_r(\theta^*) \;=\; \theta^* \;=\; \bar{\theta}(\theta^*) \;<\; \bar{\theta}(0) \;<\; \frac{\delta}{1-\delta}E(\theta) \;=\; \bar{\theta}_r(0) \tag{12}$$

Assuming (11), we now move on to the optimality of the customer's behavior in $S_r = \langle(\theta^{(0)}, \theta^{(0)}), \theta^{(1)}, \cdots\rangle$. As before, we provide a specification of behavior on off the equilibrium paths following customer's deviations, by extending the one described in Section 3 for p.a.e. Specifically, we retain (P2)∼(P4) for backup phases $k = 1, 2, \cdots$. In addition, we extend (P4) to cover the initial phase, and modify (P1) and (P3) to (P1') and (P3') below, respectively, to accommodate the initial phase.

(P1') If an agent is not consulted in period t of phase 0, he assumes $\langle\theta^{(1)}, \theta^{(2)}, \cdots\rangle$ passively in period t. If an agent is refused to see the other mechanic's report in period t, he assumes $\langle\theta^{(1)}, \theta^{(2)}, \cdots\rangle$ actively in period $t+1$.

(P3') 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 $\langle\theta^{(1)}, \theta^{(2)}, \cdots\rangle$ actively in period $t+1$. If the other mechanic, B, is not consulted in period $t+1$, he assumes $\langle\theta^{(1)}, \theta^{(2)}, \cdots\rangle$ passively in period $t+1$.

The customer has no incentive to deviate in hiring decision due to (P4), as explained 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 maneuvre a transition to phase 1 to enjoy more reliable reports, specifically by refusing the report to one mechanic and consulting only him in the next period (see (P1')).

Combining with $\theta^{(1)} \geq \theta^{(2)} \geq \cdots$, property (7) of p.a.e, again we find (7) as necessary and sufficient for optimality of the customer's behavior.

Lemma 8: *Given a rivalry sequence* $\langle(\theta^{(0)}, \theta^{(0)}), \theta^{(1)}, \cdots\rangle$, *augmented by off the equilibrium behavior as specified in (P1'), (P2), (P3), (P3') and (P4), the customer's behavior is a best response in each phase if and only if (3) holds for* $k \geq 1$, *and (12) and (7) hold, or equivalently, if and only if*

$$\theta^{(1)} \leq \theta^{(0)} \leq \bar{\theta}_r(\theta^{(1)}) \quad \textit{and} \quad \theta^{(k+1)} \leq \theta^{(k)} \leq \bar{\theta}_r(\theta^{(k+1)}) \quad \forall k \geq 1 \tag{13}$$

Since $\bar{\theta}_r(\cdot)$ is a decreasing function, the set of effective trust levels sustainable in r.a.e. is $[0, \bar{\theta}_r(0)]$. The r.a.e. with the maximum trust level is $\langle(\bar{\theta}_r(0), \bar{\theta}_r(0)), 0, 0, \cdots\rangle$.

It is worth noting from (12) that the maximum effective trust level, $\bar{\theta}_r(0)$, of r.a.e. is higher than that of p.a.e, $\bar{\theta}(0)$. 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 some period in *any* equilibrium, including non-"stationary" ones and those in which mechanics may not report with a trust level. The result is stated in Lemma 9 and is proved in Appendix B. Given an equilibrium, we say that there is *truthful revelation* for θ_t in period t if the customer hires mechanic A (B) if $\tau_t =$ A (B).

Lemma 9: *Fix an arbitrary equilibrium. If there is truthful revelation for* θ_t *in period* t, *then* $\theta_t \leq \frac{\delta}{1-\delta}E(\theta)$.

Now we apply the credibility argument of the backup phases that has been developed in the previous section. By the same spirit, we define an r.a.e. S_r to be *recursively credible* if the truncation $S_r^{(1)}$ is a recursively credible p.a.e. and there does not exist a recursively credible p.a.e. that overrides $S_r^{(1)}$. Since $S^* = \langle\theta^*, \theta^*, \cdots\rangle$ is the unique recursively credible p.a.e, an r.a.e. S_r is recursively credible if and only if $S_r^{(1)} = S^*$. Recall $\theta^{(1)} \leq \theta^{(0)} \leq \bar{\theta}_r(\theta^{(1)})$ from (13). Since $\theta^{(1)} = \theta^*$, the fixed point of $\bar{\theta}_r(\cdot)$, it is immediate to verify that $S_r^* = \langle(\theta^*, \theta^*), \theta^*, \theta^*, \cdots\rangle$ is the unique r.a.e that is recursively credible. The next theorem summarizes the findings for r.a.e.

Theorem 10: *A rivalry sequence* $\langle(\theta^{(0)}, \theta^{(0)}), \theta^{(1)}, \cdots\rangle$, *augmented by off the equilibrium behavior as specified in (P1'), (P2), (P3), (P3') and (P4), constitutes an r.a.e. if and only if (13) holds. The set of trust levels sustainable in r.a.e. is* $[0, \frac{\delta}{1-\delta}E(\theta)]$. *The sequence* $S_r^* = \langle(\theta^*, \theta^*), \theta^*, \theta^*, \cdots\rangle$ *is the unique recursively credible r.a.e.*

Finally, we show that the equilibria considered in this and previous sections effectively cover all equilibria, in the sense that for any equilibrium there exists a p.a.e. or an r.a.e. with the same consumer's and mechanics' surpluses. Consider an arbitrary equilibrium: this may not be stationary and the mechanics may not report with a trust level. In each period t along this equilibrium, the set of θ_t for truthful revelation is a subset of $[0, \bar{\theta}_r(0)]$ by Lemma

9.[18] Therefore, the expected consumer's surplus is lower than that in the "optimal" r.a.e., $\langle(\bar{\theta}_r(0), \bar{\theta}_r(0)), 0, 0, \cdots\rangle$. The total expected mechanics' surplus is the same $(\frac{1}{1-\delta}E(\theta))$ in all equilibria (only the division between two mechanics is different), because one of the mechanics provides the service in every period. Therefore, the social surplus is higher in the optimal r.a.e. than in the one arbitrarily chosen above. In fact, by selecting the initial trust level carefully, we can find an r.a.e. with the same total social surplus as the latter.

5. Extension to More Experts

We extend the analysis to the cases that 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 $\tau_t(= A, B, \cdots, N)$. There being a larger number of experts due to finer differentiation, the degree of rivalry among them are potentially higher. Our main concern is its effects on the sustainable trust level.

5.1. Primary agency equilibrium with N agents

The concept of p.a.e. 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, we denote such successive phases by a sequence $\langle\theta^{(0)}, \theta^{(1)}, \cdots\rangle$. The only difference from the two experts case analyzed in Section 3 is that each agent gets to provide the service with probability $\frac{1}{N}$ 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 accommodates this difference: each agent's behavior is a best repsonse in phase k if and only if

$$\frac{\delta}{1-\delta}\Big(\int_{\theta^{(k)}}^{\infty}\theta dF + \frac{1}{N}\int_0^{\theta^{(k)}}\theta dF - \frac{1}{N}\int_0^{\theta^{(k+1)}}\theta dF\Big) - \theta^{(k)} \;\geq\; 0 \tag{14}$$

Defining $\bar{\theta}^N(\theta^{(k+1)})$ to be the value of $\theta^{(k)}$ at which (13) is satisfied tightly, we conclude that the agents' behavior is optimal if and only if

$$0 \;\leq\; \theta^{(k)} \;\leq\; \bar{\theta}^N(\theta^{(k+1)}) \qquad \forall k = 0, 1, 2, \cdots \tag{15}$$

A specification of off the equilibrium paths is needed for optimality of the customer's behavior, which is a straightforward modification of the postulates discussed in Section 3. We omit the details here because they are a routine exercise. The main result to be

[18] In principle, one can imagine the possibility that mechanics are "partially truthful" for θ_t in the sense that they report truthfully with a probability less than 1. (However, such an equilibrium may not exist.) In fact, by an argument analogous to the proof of Lemma 9 (Appendix B), one can show that $\frac{\delta}{1-\delta}E(\theta)$ is an upper bound for partially truthful revelation in any equilibrium, too. Hence, the subsequent argument is robust to this possibility.

stated is that, as before, inequality (7) is necessary and sufficient for optimality of the customer for both p.a.e.'s and n-rivalry agency equilibria to be discussed in Section 5.2. This is because, if $\theta^{(k+1)} > \theta^{(k)}$, the customer would maneuvre a transition to phase $k+1$ to enjoy a higher level of honesty.

Therefore, a p.a.e. with N experts is characterized by (7) and (15), or equivalently, by

$$\theta^{(k+1)} \ \leq \ \theta^{(k)} \ \leq \ \bar{\theta}^N(\theta^{(k+1)}) \qquad \forall k = 0, 1, 2, \cdots \tag{16}$$

We discuss some properties of $\bar{\theta}^N(\cdot)$. It is easy to see that it 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 (14): the two terms with coefficient $\frac{1}{N}$ cancel each other out when $\theta^{(k)} = \theta^{(k+1)}$. It follows from (16) that the range of possible initial trust level is $[0, \bar{\theta}^N(0)]$, and the range of possible backup trust levels is $[0, \theta^*]$.

It is straightforward to verify that $\bar{\theta}^N(\theta) > \bar{\theta}^{N+1}(\theta)$ for $\theta \in [0, \theta^*]$. The intuition is as follows. Since each agent provides the service less frequently in the trusted range for a larger N, the lower is the expected future payoff after cheating, which discourages cheating; at the same time, the expected future payoff from staying faithful is also lower, which encourages cheating. If $\theta \in [0, \theta^*]$, the backup trust level is low enough for the latter effect to dominate the former. So, we have $\bar{\theta}^2(0) > \bar{\theta}^3(0) > \cdots$, that is, the honesty level that p.a.e. can sustain deteriorates as there are more experts.

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

Theorem 11: *Suppose there are N experts. A sequence $\langle \theta^{(0)}, \theta^{(1)}, \cdots \rangle$ constitutes a p.a.e. if and only if (16) holds. The set of trust levels sustainable in p.a.e. is $[0, \bar{\theta}^N(0)]$ where $\bar{\theta}^N(0)$ decreases in N. The sequence $S^* = \langle \theta^*, \theta^*, \cdots \rangle$ is the unique recursively credible p.a.e. for each $N = 2, 3, \cdots$.*

Corollary 12: *As there are more experts, i) the maximum honesty level sustainable by a p.a.e. decreases, but ii) the honesty level of recursively credible p.a.e. stays the same at θ^*.*

5.2 Collusion and n-rivalry agency equilibrium

In the initial phase of an n-rivalry agency equilibrium (n-r.a.e, hereafter) 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, we focus on "symmetric" equilibria in which the n agents report with a common, initial trust level $\theta^{(0)}$; the customer, then, hires the n agents for repair service with even probability, $\frac{1}{n}$, 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 say that the customer *trusts* the panel if she behaves as 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 non-cheaters (i.e, $(n-1)$-rivalry), or she may lose interest in r.a.e. altogether and resort to a p.a.e. However, what determines the incentives to cheat and consequently, the sustainable honesty level, is the backup trust level that will prevail in the punishment phase.

In line with the previous sections, we first find the maximum level of honesty sustainable by an n-r.a.e. without restrictions on the backup phases: since lower backup trust level induces higher initial trust level, we do this by setting the backup trust level at 0, or more specifically, we set the p.a.e. $\langle 0, 0, \cdots \rangle$ as the sequence of backup phases. Then, we impose the credibility criterion to find the n-r.a.e. that is recursively credible.

Consider a panel member in the initial phase of an n-r.a.e. 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 are 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, we consider collusions by agents who may agree to misreport in a coordinated way to mislead the customer's decision and to split the proceeds evenly among themselves.[19]

To find the maximum honesty level sustainable by n-r.a.e, let $\langle 0, 0, \cdots \rangle$ be the sequence of backup phases. Then, optimality in non-initial phases is automatic and we focus on the initial phase. Consider a panel member, say mechanic A, of an n-r.a.e. with trust level $\theta^{(0)}$. In the case that $\theta_t < \theta^{(0)}$ and $\tau_t \neq A$, for an effective deviation he needs to form a collusion consisting of at least $n/2$ members. Since a larger collusion reduces his share of proceeds from deviation, 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 $\frac{1}{2}$ chance of success (because the customer will evenly randomize the right mechanic and the mechanic recommended by the collusion), whilst a collusion of $(n/2)+1$ members is assured of success but each member's share is smaller. We examine even-numbered r.a.e.'s first and then verify that odd-numbered r.a.e.'s perform worse.

Before proceeding, two comments are in order on the credibility of recommendations made by collusions. To see the first point, suppose the customer faces two mechanics recommended by the panel, one of whom is a panel member and the other is not. Then, she would reason that the former is recommended by a collusion, because recommending a

[19] Because of symmetry, collusions are easier to form when the proceeds are split evenly. However, we do not discuss the issue of enforceability of collusion agreement because it is beyond the purposes of this paper.

mechanic outside the panel is not an efficient collusion: what the collusion members share would be less than the whole proceeds because the mechanic who performed the service would demand his share, too. If instead they deviated by recommending one collusion member when τ_t is the specialty of one panel member (but not a collusion member), then the customer would not have inferred the collusive recommendation and the collusion would have done better because they share the proceeds only among themselves. The collusive deviations dealt with in this section are of this sort.

Theoretically more interesting is the possibility of inferring the collusive recommendation from the split of panel members between two recommended mechanics. If, for instance, mechanics A and B are recommended by two and three panel members, respectively, at first instance the customer might be inclined to infer mechanic B as the collusive one because a minority collusion does not make sense. But, such inference would be backfired by rendering two-member collusions indeed effective and thereby, enhancing cheating incentives. The majority rule that we adopted for hiring decision in case of disagreement, is the one that minimizes the cheating incentives.[20]

Returning to the main task, consider an n-r.a.e. where n is even. 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 $\frac{1}{2}$, 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 (same when $n=2$).[21] Since the future expected payoff of a cheater is 0 in any case, if it is not profitable to form a collusion of $(n/2)+1$ members, neither is to form any other collusion. We formulate this condition below.

From above we 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

[20] In this discussion we implicitly assumed that the customer hires among the recommended mechanics. To reduce cheating incentives by lowering the success rate, she may stretch randomization to include mechanics who have not been recommended by the panel. But, this generates further complications. For example, consider randomizing among all N mechanics in case of any disagreement, which seemingly reduces the success rate most. To see that this rule is not sensible, consider a panel member, say A, who has not been recommended. For him still to be included in the randomization rule, the customer must believe that he belonged to a collusion and recommended someone else even when he turned out to be the right mechanic. This is absurd because if he were the right mechanic, by recommending himself he would have had the same expected payoff in the current period as sticking with the collusion (because collusion members will share the proceeds whenever one of them gets hired), but without future punishment.

Hence, the most effective hiring rule that includes non-recommended mechanics is to randomize among the most recommended mechanic and all non-panel members in case of disagreement. But, this opens up the problem of bargaining over sidepayments between the collusion and non-panel members. Since the latter have weak threat points, the bargaining power appears to be on the collusion side. If this is the case, the analysis in the paper stays vaild.

In addition, it is delicate to justify a hiring rule that randomizes beyond the most recommended mechanics: the limiting behavior that generates a compatible "consistent assessment" needs be highly concerted among experts to warrant such a hiring rule. On the other hand, the majority rule is justified by the limit of a sequence of simple, completely mixed strategies of experts, namely, making small, symmetric mistakes in their reporting.

[21] When $n=2$, a collusion of both panel members is absurd. But, mathematically it is equally profitable as the more sensible, one-member collusion and so the analysis is unaffected.

is maintained for all $\theta_t \leq \theta^{(0)}$, is

$$\frac{\delta}{1-\delta}\Big(\frac{1}{n}\int_{\theta^{(0)}}^{\infty} \theta dF + \frac{1}{N}\int_{0}^{\theta^{(0)}} \theta dF\Big) - \frac{2}{n+2}\,\theta^{(0)} \;\geq\; 0 \tag{17}$$

Define $\bar{\theta}_n^N(0)$ to be the value of $\theta^{(0)}$ at which (17) is satisfied tightly (the argument 0 in $\bar{\theta}_n^N(0)$ signifies the backup trust level): $\bar{\theta}_n^N(0)$ is the highest $\theta^{(0)}$ subject to (17) because LHS of (17) is decreasing in $\theta^{(0)}$. Since the optimality of the customer is warranted by (7) as mentioned earlier, $\bar{\theta}_n^N(0)$ is the maximum honesty level sustainable by n-r.a.e. The value $\bar{\theta}_r(0)$ in Section 4 is the special case that $N = n = 2$.

Lemma 13: *Suppose $N \geq 2$ and n is an even number between 2 and N.*
(a) *If $N > 2$, then $\bar{\theta}_n^N(0) < \bar{\theta}_r(0)$.*
(b) *If $N < N'$, then $\bar{\theta}_n^N(0) > \bar{\theta}_n^{N'}(0)$.*

Proof: Part (b) is immediate from (17): $\theta^{(0)} = \bar{\theta}_n^N(0)$ violates (17) for N' and, therefore, part (b) follows.

Part (a): Since $\bar{\theta}_r(0)$ solves (17) tightly when $N = n = 2$, we have

$$\frac{\delta}{1-\delta}\Big(\frac{2}{n+2}\int_{\bar{\theta}_r(0)}^{\infty} \theta dF + \frac{2}{n+2}\int_{0}^{\bar{\theta}_r(0)} \theta dF\Big) = \frac{2}{n+2}\,\bar{\theta}_r(0)$$

Since $\frac{1}{N} \leq \frac{1}{n} \leq \frac{2}{n+2}$ and at most one inequality holds tightly, it follows that

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

which violates (17). Therefore, part (a) follows. QED

Part (b) is intuitively clear because the reward for being faithful is smaller when N is large. The intuition for part (a) is that when n and N increase from 2, the reward for being faithful decreases more than the gain from (collusive) deviation does, because the proceeds for the latter are shared by fewer agents.

However, the effect of increasing the size of panel is not clear-cut. That is, $\bar{\theta}_{n+2}^N(0)$ may be higher or lower than $\bar{\theta}_n^N(0)$ for a fixed N. This is because although a larger panel size reduces the reward for being faithful by a bigger factor, it affects only part of it, namely, the part corresponding to the trusted range of θ.

Next, consider odd-numbered n-r.a.e. As 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 $(n-1)$-r.a.e. discussed above and, therefore, the expected gain from collusive deviation is the same. But, the reward from being faithful is higher in $(n-1)$-r.a.e. 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 $(n-1)$-r.a.e. So,

$$\bar{\theta}^N_{n-1}(0) \; > \; \bar{\theta}^N_n(0) \;\; \text{if } n \text{ is odd.}$$

Combining with Lemma 13, we have the following conclusion.

Theorem 14: *As there are more experts due to finer differentiation, the maximum trust level sustainable by n-r.a.e. $(2 \leq n \leq N)$ strictly deteriorates. With a given number of experts, the panel that sustains the maximum honesty consists of an even number of members, but the exact size is ambiguous.*

Finally, we extend the credibility criterion to n-r.a.e.'s. In fact, we generalize Definition 1 in Section 3 to cover all the cases and equilibria considered in this paper. Specifically, for $N \geq n$ and $n = 1, 2, \cdots$, an n-r.a.e. is an infinite sequence $S_{n/N}$ of phases, each phase characterized by a panel (contingent on the 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-r.a.e. where m is a number between 1 and N: in particular, $m > n$ is possible. However, the cheaters in a 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-r.a.e.'s includes all p.a.e.'s and more.

In earlier analyses of optimality of the agents, we assumed that the customer would patronize a trusted agent in each backup phase. In "more general" n-r.a.e. described in the previous paragraph, the backup phase may be served by a panel. However, the players' incentives in the current phase are determined by the trust level of the subsequent backup panel but not by 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-r.a.e.'s. In particular, the functions $\bar{\theta}(\cdot), \bar{\theta}_r(\cdot)$ and $\bar{\theta}^N_n$ are valid.

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

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

(b) Let $k > 1$ and assume that round-k' credible n-r.a.e. has been defined for all $k' < k$ and all $n = 1, \cdots, N$. Then, an n-r.a.e. $S_{n/N}$ is *round-k credible* if
  - i) $S^{(1)}_{n/N}$ is round-$(k-1)$ credible, and
  - ii) there does not exist a *round-k credible* m-r.a.e. that overrides $S^{(1)}_{n/N}$.

(c) An n-r.a.e. $S_{n/N}$ is *recursively credible* if it is round-k credible for all $k = 1, 2, \cdots$.

This is a straightforward generalization of Definition 1 in Section 3. Consequently, an argument exactly analogous to the one in Section 3 allows us to deduce that candidates for recursively credible n-r.a.e. have the same trust levels, between 0 and θ^*, for all phases.

Now, we verify that an n-r.a.e. $S_{n/N}$ is recursively credible if and only if θ^* is the common trust level for all phases. If θ^* is the common trust level, then $S^{(1)}_{n/N}$ is not overriden by a round-k credible m-r.a.e. because θ^* is the maximum possible initial trust level for any such r.a.e. Since this is true for all k, $S_{n/N}$ is recursively credible. We already found such equilibria for $n = 1$ and 2: $S^* = \langle\theta^*, \theta^*, \cdots\rangle$ is a recursively credible p.a.e. and $S^*_r = \langle(\theta^*, \theta^*), \theta^*, \cdots\rangle$ is a recursively credible 2-r.a.e. In fact, 1- and 2-r.a.e's consisting of phases with one- or two-member panels with trust level θ^* are all recursively credible.

However, it turns out that there is no n-r.a.e. 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 $n \geq 3$. To see this, calculate the condition for there being no incentive for a panel member to cheat in the initial phase:

$$\frac{\delta}{1-\delta}\Big(\frac{1}{n}\int_{\theta^{(0)}}^{\infty}\theta dF + \frac{1}{N}\int_0^{\theta^{(0)}}\theta dF - \frac{1}{N}\int_0^{\theta^*}\theta dF\Big) - \frac{2}{n+2}\theta^{(0)} \;\geq\; 0 \tag{18}$$

when n is even; when n is odd the coefficient of the last term is $\frac{2}{n+1}$. If $n = 2$, the value of LHS of (18) is one half of the value of LHS of (14) when $\theta^{(0)} = \theta^*$: since the latter is 0, so is the former. Hence, $\bar{\theta}^N_2(\theta^*) = \theta^*$ which verifies that S^*_r above is indeed a 2-r.a.e. Compare the LHS of (18) when $n \geq 3$ with the case $n = 2$ for $\theta^{(0)} = \theta^*$:

$$\frac{\delta}{1-\delta}\Big(\frac{1}{n}\int_{\theta^*}^{\infty}\theta dF\Big) - \frac{2}{n+2}\theta^* \;<\; \frac{\delta}{1-\delta}\Big(\frac{1}{2}\int_{\theta^*}^{\infty}\theta dF\Big) - \frac{2}{2+2}\theta^* \;=\; 0 \tag{19}$$

when $n \geq 3$, violating (18). This implies that the maximum initial trust level sustainable, given a backup trust level of θ^*, is strictly lower than θ^*. When $n \geq 3$, therefore, it is not possible for an n-r.a.e. to have θ^* as the trust level for all phases.

Theorem 15: *Suppose there are $N \geq 3$ differentiated experts. Recursively credible n-r.a.e.'s exist for $n = 1, 2$: they have the same trust level, θ^*, for all phases, and each phase has either one- or two-member panel. For $n \geq 3$, a recursively credible n-r.a.e. does not exist.*

Appendix A: Proof of Lemma 6

First, find mechanic B's optimality condition in the initial phase of $\langle(\theta_A, \theta_B), \theta^{(1)}, \theta^{(2)}, \cdots\rangle$. As explained earlier, cheating is feasible only when $\theta_t < \theta_B$, and other things being equal, incentive to cheat is greater when $\tau_t = A$ that when $\tau_t = B$. Hence, consider mechanic B in period t of the initial phase, who examined the car and learned $\theta_t < \theta_B$ and $\tau_t = A$. If he follows the supposed strategy throughout and so the initial phase is maintained, the discounted sum of his expected payoff stream is

$$\frac{\delta}{1-\delta}\Big((1-p)\int_{\theta_A}^{\infty}\theta dF + \frac{1}{2}\int_0^{\theta_A}\theta dF\Big) \tag{20}$$

because he gets 0 now and, in each of future period t', he will provide the repair service with a probability $(1-p)$ if $\theta_{t'} \geq \theta_A$ and he will do so for the case $\tau_{t'} = B$ if $\theta_{t'} < \theta_A$. On the other hand, if he cheats in this period, he gets $\theta_t\,(< \theta_B)$ with a probability of $(1-q)$ now and the first backup phase prevails from next period onwards. This generates a discounted sum of

$$(1-q)\theta_t + \frac{\delta}{2(1-\delta)} \int_0^{\theta^{(1)}} \theta dF \tag{21}$$

So, it would never be profitable for mechanic B to cheat if and only if (20) is at least as large as (21) for all $\theta_t < \theta_B$. Since (21) is increasing in θ_t, this is equivalent to

$$\frac{\delta}{1-\delta}\Big((1-p)\int_{\theta_A}^{\infty} \theta dF + \frac{1}{2}\int_0^{\theta_A} \theta dF - \frac{1}{2}\int_0^{\theta^{(1)}} \theta dF\Big) - (1-q)\theta_B \;\geq\; 0 \tag{22}$$

Next, we find mechanic A's optimality. 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

$$\frac{\delta}{1-\delta}\Big(p\int_{\theta_A}^{\infty} \theta dF + \frac{1}{2}\int_0^{\theta_A} \theta dF - \frac{1}{2}\int_0^{\theta^{(1)}} \theta dF\Big) - q\,\theta_B \;\geq\; 0 \tag{23}$$

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

$$\frac{\delta}{1-\delta}\Big(p\int_{\theta_A}^{\infty} \theta dF + \frac{1}{2}\int_0^{\theta_A} \theta dF - \frac{1}{2}\int_0^{\theta^{(1)}} \theta dF\Big) - \theta_A \;\geq\; 0 \tag{24}$$

It is straightforward that (24) implies (23) because $\theta_A \geq \theta_B$.

So far we have characterized the optimality condition of the agents with (22) and (24). 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, inequality (24) drops out as an optimality condition. In this case, by symmetry, mechanic A's optimality condition coincides with (22) where $(1-p)$ and $(1-q)$ are replaced by p and q, respectively:

$$\frac{\delta}{1-\delta}\Big(p\int_{\theta_A}^{\infty} \theta dF + \frac{1}{2}\int_0^{\theta_A} \theta dF - \frac{1}{2}\int_0^{\theta^{(1)}} \theta dF\Big) - q\theta_B \;\geq\; 0 \tag{25}$$

Adding (22) and (25) side by side and taking a half of both sides (remember $\theta_A = \theta_B$), we get

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

which coincides with (22) and (25) for the case $p = q = \frac{1}{2}$. 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 $(\theta_A, \hat{\theta}_B = \theta_A, \frac{1}{2}, \frac{1}{2})$. This completes the proof.

Appendix B: Proof of Lemma 9

Suppose that truthful revelation occurs for some θ_t in a particular period t. Then, it must be the case that at least one mechanic, say A, reports truthfully for θ_t: 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 can not hire the right mechanic with certainty, contradicting truthful revelation.

Let m^A and m^B denote the messages that mechanic A sends when $\tau_t = A$ and B, respectively. Let n^A, n^B and n^C denote the messages that mechanic B may send only when $\tau_t = A$, only when $\tau_t = B$, and in either contingencies, respectively. Because the right mechanic is hired all the time for θ_t in the equilibrium, the customer's response to the received message pair must satisfy:

(a) 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 the equilibrium), she may randomize:

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

Let V_s^A denote the expected payoff in period s for mechanic A in the equilibrium. (Since the equilibrium under question need not be stationary, the expected payoff is period-dependent.) Then, $E(\theta) - V_s^A$ denotes the expected payoff in period s for mechanic B because the service is provided by one of the two mechanics in each period.

Suppose $\tau_t = B$. The expected current payoff for mechanic A from cheating is at least $r\theta_t$. (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):

$$\sum_{s=t+1}^{\infty} \delta^{s-t} V_s^A \;\geq\; r\theta_t \tag{26}$$

Next, suppose $\tau_t = A$. The expected current payoff for mechanic B from cheating, i.e, sending n^B, is $(1 - r)\theta_t$. Since cheating does not occur in the equilibrium, it must be the case that the equilibrium expected payoff is larger:

$$\sum_{s=t+1}^{\infty} \delta^{s-t} (E(\theta) - V_s^A) \;\geq\; (1 - r)\theta_t \tag{27}$$

Adding (26) and (27) side by side, we prove

$$\frac{\delta}{1-\delta} E(\theta) \;\geq\; \theta_t \,.$$

References

Austen-Smith, D. (1990), "Information transmission in debate," *Amer. J. Pol. Sci.* **34**, 124-152.

Benabou, R. and Laroque, G. (1992), "Using privileged information to manipulate markets: Insiders, gurus and credibility," *Quart. J. Econ.* **107**, 921-958.

Bernheim, B. D., Peleg, B. and Whinston, M. (1987), "Coalition-proof Nash equilibria I. Concepts," *J. Econ. Theory* **42**, 1-12.

Blume, A. (1994), "Equilibrium refinements in sender-receiver games," *J. Econ. Theory* **64**, 66-77.

Crawford, V. P. and Sobel, J. (1982), "Strategic Information Transmission," *Econometrica* **50**, 1431-1451.

Darby, M. and Karni, E. (1973), "Free competition and the optimal amount of fraud," *J. Law Econ.* **16**, 67-88.

Dewatripont, M. and Tirole, J. (1999), "Advocates," *J. Pol. Econ.* **107**, 1-39.

Gilligan, T. and Krehbiel, K. (1989), "Asymmetric information and legislative rules with a heterogeneous committee," *Amer. J. Pol. Sci.* **33**, 459-490.

Kim, J.-Y. (1996), "Cheap talk and reputation in repeated pretrial negotiation," *Rand J. Econ.* **27**, 787-802.

Kreps, D. M. and Wilson, R. (1982), "Sequential equilibria," *Econometrica* **50**, 863-894.

Klein, B. and Leffler, K. (1981), "The role of market forces in assuring contractual performances," *J. Pol. Econ.* **89**, 615-641.

Krishna, V. and Morgan, J. (1999), "A model of expertise," mimeo, Penn State University and Princeton University.

Lipman, B. and Seppi, D. (1995), "Robust inference in communication games with partial provability," *J. Econ. Theory* **66**, 370-405.

Morris, S. (1998), "An instrumental theory of political correctness," mimeo, University of Pennsylvania and Northwestern University.

Ottaviani, M. and Sorensen, P. (1999a), "Information aggregation in debate: who should speak first?," mimeo, University College London and University of Copenhagen.

Ottaviani, M. and Sorensen, P. (1999b), "Professional advice," mimeo, University College London and University of Copenhagen.

Park, I.-U. (1997), "Generic finiteness of equilibrium outcome distributions for sender-receiver cheap-talk games," *J. Econ. Theory* **76**, 431-448.

Pesendorfer, W. and Wolinsky, A. (1998), "Second opinion and price competition: Inefficiency in the market for expert advice," CMSEMS DP 1229, Northwestern.

Pitchik, C. and Schotter, A. (1987), "Honesty in a model of strategic information transmission," *Amer. Econ. Rev.* **77**, 1032-1036.

Satterthwaite, M. A. (1979), "Consumer information, equilibrium industry price, and the number of sellers," *Bell J. Econ.* **10**, 483-502.

Shapiro, C. (1983), "Premiums for high quality products as returns to reputations," *Quart. J. Econ.* **98**, 659-679.

Shin, H. S. (1994), "The burden of proof in a game of persuation," *J. Econ. Theory* **64**, 253-264.

Sobel, J. (1985), "A theory of credibility," *Rev. Econ. Stud.* **52**, 557-573.

Taylor, C. (1995), "The economics of breakdowns, checkups, and cures," *J. Pol. Econ.* **103**, 53-74.

Wolinsky, A. (1993), "Competition in a market for informed experts' services," *Rand J. Econ.* **24**, 380-398.