

STRATEGIC REFERRALS AMONG EXPERTS*

YI CHEN[†] THOMAS JUNGBAUER[†] MARK SATTERTHWAITE[§]

March 1, 2025

ABSTRACT. We study a dynamic model of referrals among experts that are horizontally specialized in treating heterogeneous problems. Demand for an expert's services increases in their reputation, a decaying record of their historic success rate in solving problems. Experts refer mismatched problems, trading off today's revenue with a higher demand tomorrow. In the unique steady state of Markov equilibria, the market generally neither supports efficient referrals nor optimal specialization, a problem potentially exacerbated by the adoption of AI. Our results suggest that regulation should allow for referral alliances and partnerships among small numbers of experts with complementary skills.

Keywords: Referrals, Reputation, Dynamic Games, Markov Equilibrium, AI

JEL Codes: C73, D83, I11, J44, L84

*We are grateful for helpful comments by Ian Ball, Arjada Bardhi, V. Bhaskar, William Fuchs, Drew Fudenberg, George Georgiadis, Michael Greinecker, Marina Halac, Wei He, Mitchell Hoffman, Justin Johnson, Andrew Koh, Fei Li, Jin Li, Niko Matouschek, Mallesh Pai, Michael Powell, Marcel Preuss, Klaus Ritzberger, Vasiliki Skreta, Curtis Taylor, Michael Waldman and unnamed participants of various conferences and seminars. Earlier versions of the paper were presented as "Referral Behavior when Matching is Important." Finally, we would like to acknowledge excellent research assistance by Si Zuo. All remaining errors are our own.

[†]SC Johnson College of Business, Cornell University. Email: yi.chen@cornell.edu.

[‡]SC Johnson College of Business, Cornell University. Email: jungbauer@cornell.edu.

[§]Kellogg School of Management, Northwestern University.

1 Introduction

Matching problems with expertise within and across firms is essential for the functioning of the economy. When problems are complex, their owners are frequently unaware of their nature, and costly upfront diagnosis by experts is required. The initial allocation of problems across heterogeneous experts is therefore often independent of experts' specialization, but relies on information about their past success in solving problems. Prospective clients, for example, prefer to consult lawyers with better track records or doctors with higher patient satisfaction scores, while firms are more likely to assign projects to managers who got the job done in the past. The services of experts in these scenarios are known as "reputation goods" (Satterthwaite, 1979).¹ In order to strengthen their reputation—thereby increasing demand for their services—experts may find it in their best interest to forego compensation for treating a problem themselves, but refer it to another expert who is a better fit.

In this paper, we introduce a dynamic model of strategic referrals among specialized experts to shed light on how experts' reputational concerns shape market outcomes. We find that competition for problems among individual experts does in general not result in an efficient allocation of problems across experts. What is more, instead of sharpening their expertise, experts may want to invest in their abilities outside of their specialty—to the detriment of the economy. This issue may be exacerbated by the adoption of Artificial Intelligence (AI) in the workplace, often increasing professionals' productivity outside their main specialty. Also, market efficiency increases in the experts' compensation for diagnosing problems relative to treatment, potentially discouraging them from specializing in the first place. We show that referral alliances that set referral prices may alleviate these issues, though possibly at the expense of some experts. Pareto-efficient market outcomes generally require revenue sharing beyond referrals, i.e., some form of partnership contract.

Our model features an economy with a mass of experts of two different types competing for problems in continuous time. At each instant, every expert receives a flow of problems, their instantaneous *demand*. Each problem's *category* matches one of the

¹A *reputation good* is (1) differentiated between sellers, (2) product quality is consumer specific (3) and initially uncertain, (4) important, and thus (5) worth considerable information finding effort.

expert types, where an expert has a higher success rate in treating a problem in their category, i.e., one that matches their type. Since determining a problem's category requires costly *diagnosis* by an expert, an expert's initial demand is unsorted, i.e., the mix of problems they receive is independent of their type. The expert initially diagnoses their problems, and then decides for each of them whether to *treat* or to *refer* the problem. Experts earn fees every time they diagnose or treat a problem.

Despite the foregone treatment fees, experts still may want to refer problems to other experts due to reputational concerns. Specifically, an expert can *solve* a problem by either treating it successfully themselves, or by referring it to another expert who then treats it successfully. After treatment outcomes realize, the expert's *record* updates.² This public record is specified by its *thickness* and the expert's *reputation*. Reputation—a decaying weighted average of historic success in solving problems—is directly payoff-relevant. In particular, the demand for an expert's services increases in their reputation but decreases in some form of aggregate reputation, perhaps the market average. The record's thickness—a decaying sum of problems diagnosed by the expert—does not affect demand for the expert's services. It, however, determines the “inertia” of reputation and is crucial for how easily the expert's current success rate in solving problems affects reputation.

As such, we solve a dynamic game among infinitely many experts of two types, where each expert conditions their action on all experts' two-dimensional records. We focus on Markov equilibria of the referral game, in the sense that experts make their referral decisions solely based on their own and the aggregate reputations. Using a recursive method on each expert, we establish the existence of a unique steady state of Markov equilibria. Intuitively, the uniqueness stems from a negative feedback loop of the record system. Suppose an expert aggressively refers problems to more apt experts. As a result, the thickness of this expert's record grows, making their reputation more “inert” to their current success rate. This reduces their incentives to refer, until reaching a tipping point at which the reputational benefit from referral is no longer worth losing the treatment fee. As a result, the expert ceases to refer, their demand decreases, and their record again loses its thickness. It is this

²We think of the expert's record as the entirety of word-of-mouth, online or performance reviews that indicate the expert's eventual success rate in solving problems they received.

negative feedback loop with endogenous control—reminiscent of reputation building and milking cycles in [Liu \(2011\)](#)—giving rise to a unique steady state.

Invoking our theoretical results, we find that reputation goods markets do in general not induce the first-best outcome in their steady state. In other words, some problems are not treated by the most apt experts in the economy. This is because the main trade off an expert faces—refer a mismatched problem to increase tomorrow’s demand through a higher level of reputation but lose today’s treatment fee—is not powerful enough to discipline the expert’s referral behavior. The resulting mismatch of experts and problems is less of a concern the more specialized experts in the economy are. A larger *leading* ability does not only directly affect success rates when treating problems but also compels other experts to refer more often. Experts themselves, however, often prefer to increase their *lagging* ability instead, i.e., their success rate in treating mismatched problems. This lowers overall welfare as experts experience a diminished incentive to refer to a better suited competitor.

A timely application of experts increasing their lagging ability is the adoption of Artificial Intelligence (AI). See, among others, [Brynjolfsson, Li, and Raymond \(2023\)](#) for evidence suggesting Generative AI may substantially increase lagging but not leading abilities in the workplace. As a result, a comprehensive adoption of AI may nudge professionals to refer problems outside their direct expertise less often. In this scenario, the welfare effect of AI is unambiguously negative.

Efficient referral behavior by experts for all possible parameter combinations requires the diagnostic fee to be large relative to the treatment fee. Implementing a fee structure in which diagnostic dominates treatment income, however, gives rise to a whole slew of problems, such as a lack of incentives for experts to specialize in the first place. More generally, a larger treatment fee always leads to lower equilibrium referral and success rates in solving problems in the economy.

Finally, we examine experts’ incentives to enter referral alliances and partnerships that specify referral prices. Antitrust regulation aims to curtail horizontal mergers that increase market power. While we do not model price competition explicitly, we show the existence of an equilibrium with only small numbers of experts entering partnerships, none of which commands market power. These partnerships

are stable against joint deviations by experts and induce efficient referral behavior. Partnerships differ from referral alliances in that they also admit revenue sharing beyond referrals. While referral alliances may be sufficient to induce efficiency, the resulting welfare gains may come at the cost of some experts. Partnerships, on the other hand, allow for Pareto-efficient outcomes through income redistribution. These strategic alliances may also resolve issues with specialization and the fee structure. Our results suggest it is crucial for regulation to allow for the formation of strategic alliances when reputational concerns matter.

While we focus on referrals and their effect on the matching of problems with experts, it is important to note that the model presented in this paper lends itself well to analyze other important reputation goods markets. For example, instead of deciding whether to refer a problem or not, a firm may decide whether to sell low- or high-quality products (see, e.g., [Varas, Marinovic, and Skrzypacz, 2020](#)), or exert different levels of effort. Finally, our approach to solving for a steady state of Markov equilibria with the intertwined state variables thickness and reputation proves intuitive, and is tractable beyond these applications.

This paper is organized as follows. The remainder of this section relates our contribution to the existing literature. Section 2 introduces the setup for our dynamic model of strategic referrals among experts. Thereafter, Section 3 details our analysis of the model and establishes both existence and uniqueness of a Markov equilibrium giving rise to a unique steady state. In Section 4 we investigate how the primitives of the economy affect equilibrium referral behavior, and thus, market outcomes. Among other analyses, we examine the effect of experts' specialization, and how specialization relates to the adoption of AI in the workplace. Section 5 discusses referral alliances and partnerships before Section 6 relates our analysis in more depth to real-world markets and discusses additional literature, while Section 7 concludes.

Related Literature. This paper relates to the literature about referrals, reputation, and the organization of knowledge. The only papers—to our knowledge—that relate reputation to referrals are [Bolton, Freixas, and Shapiro \(2007\)](#) and [Inderst and Ottaviani \(2009\)](#). In [Bolton, Freixas, and Shapiro \(2007\)](#), competing financial intermediaries can inform customers that their product is not a good match, while

[Inderst and Ottaviani \(2009\)](#) discuss firms selling through agents who enjoy limited liability, and therefore do not internalize the firm's cost from selling the wrong product to a customer. In these papers, negative client feedback manifests as an exogenous firm cost in a static environment.

[Garicano and Santos \(2004\)](#) analyze a static model of referrals between two experts that differ in productivity and production costs. Referrals are for monetary compensation, and expert behavior is generally inefficient. The inefficiency results from asymmetric information about problem severity paired with unobservable effort by experts. In our paper, the driving force behind referrals—dynamic reputational concerns—is generally insufficient to discipline referral behavior, resulting in mismatch of problems and experts. [Park \(2005\)](#) presents a dynamic cheap talk model with referrals but focuses on a single long-run consumer who punishes lying firms by withholding business. [Grassi and Ma \(2016\)](#) consider experts with comparative cost advantages in treating different problems. Organizations can overcome referral inefficiency, but at the cost of shirking.

[Garicano \(2000\)](#), [Garicano and Santos \(2004\)](#), [Fuchs and Garicano \(2010\)](#) and [Fuchs, Garicano, and Rayo \(2015\)](#) find organizations generally superior in matching expertise with unsolved problems, while markets are better in terms of utilization. While we also show that referral alliances may alleviate mismatch, efficiency in our setting does not require an organizational structure such as joint billing practices.

Finally, our approach deviates from the classic literature on reputation (see, e.g. [Fudenberg and Levine, 1999](#); [Ely and Välimäki, 2003](#)). First, in these papers, a firm's reputation is synonymous with consumer beliefs that the firm is of a certain time-invariant type. In our paper, on the other hand, reputation denotes experts' historic success rates in solving problems, a function of both their types and actions. As such, our concept is more akin to [Tadelis \(1999\)](#)—defining reputation as a firm's track record—and [Varas, Marinovic, and Skrzypacz \(2020\)](#), where reputation signals endogenous firm quality.³ Second, in our model, it is not absolute reputation that determines an expert's demand, but their reputation relative to the market.⁴

³For an excellent overview of reputation from a seller's perspective see [Bar-Isaac and Tadelis \(2008\)](#).

⁴For an early concept of aggregate reputation see [Tirole \(1996\)](#).

2 The Model

Primitives. Time $t \geq 0$ is continuous with infinite horizon. There is a unit measure of experts $i \in \mathcal{I} \equiv [0, 1]$, each either of type A or B . Let \mathcal{I}_j denote the set of experts of type $j \in \{A, B\}$, with w_j its measure. There are problems in two exclusive categories, also labeled as A and B . A problem belongs to category $j \in \{A, B\}$ with probability p_j , where $p_A + p_B = 1$. If an expert of type j treats a problem in category j' , there is a binary outcome: Success with probability $\eta_j^{j'} \in [0, 1]$ and failure otherwise. We assume that experts are horizontally specialized in that $\eta_j^j > \eta_{-j}^j$, i.e., a type A expert has an absolute advantage in treating problems in ‘their’ category A , and vice versa.⁵ We refer to the pair (η_j^j, η_j^{-j}) as a type j expert’s ability.⁶

Diagnosis, Treatment, Referral. At every instant $t \geq 0$, each expert i receives a flow f_t^i of new problems that are initially unsorted, i.e., the share of problems in category j is p_j regardless of the expert’s type. For every problem, the expert first conducts a *diagnosis* to identify its category, which—for simplicity—is perfectly accurate.⁷ If the problem’s category matches the expert’s type, then the expert *treats* it. If not, the expert can either treat it themselves, or *refer* it to an expert of the other type. Let $r_t^i \in [0, 1]$ denote the probability at time t that expert i refers a *mismatched* problem.⁸ Meanwhile, when an expert receives a referred problem, there is no more diagnosis needed but they must treat it.⁹ Finally, treatment outcomes realize.

Record. Expert i ’s time t record \mathcal{R}_t^i is a publicly observable pair (m_t^i, q_t^i) such that

$$\begin{aligned} m_t^i &\equiv m_0^i + \int_0^t f_s^i e^{-\phi(t-s)} ds, \\ q_t^i &\equiv \frac{1}{m_t^i} \left(m_0^i q_0^i + \int_0^t f_s^i \eta_s^i e^{-\phi(t-s)} ds \right), \\ \text{where } \eta_t^i &\equiv p_j \eta_j^j + p_{-j} (r_t^i \eta_{-j}^{-j} + (1 - r_t^i) \eta_j^{-j}), \text{ if } i \in \mathcal{I}_j. \end{aligned}$$

⁵Note that we discuss below why our results trivially extend to the case in which one type of expert has an absolute advantage in treating problems from both categories.

⁶We will refer to experts (and problems) j and $-j$ throughout, denoting j as an arbitrary element of $\{A, B\}$ and $-j \in \{A, B\}$ with $j \neq -j$.

⁷Incorporating flawed diagnosis does not materially affect the qualitative results in this paper.

⁸Equivalently, r_t^i denotes the proportion of mismatched problems that expert i refers in period t .

⁹As will become clear below, it is without loss of generality to assume that an expert cannot refer a problem that matches their own type.

The first component, m_t^i , denotes the *thickness* of the record, i.e., the total discounted quantity of problems expert i diagnosed by time t , with exponentially decaying weight assigned to problems in the past, and $\phi > 0$ represents the speed of decay. The second component, q_t^i , reflects expert i 's *reputation*, the weighted average rate at which expert i solves the problems they initially receive. An expert is said to solve a problem if they treat it successfully themselves or refer it to another expert who subsequently treats it successfully. The term η_t^i represents expert i 's instantaneous success rate at time t of solving problems overall, including.¹⁰ Importantly, we accredit the success of a problem to the expert who initially diagnoses it, not the one who treats it after referral. One may think of the record as online reviews or word-of-mouth recommendations indicating whether consulting an expert ultimately resulted in success or failure, with more recent cases carrying more weight.

Demand. We refer to the flow of problems to expert i in period t , f_t^i , as the instantaneous *demand* for expert i 's services. This demand $f_t^i = f(q_t^i, Q_t)$ depends on the expert's reputation q_t^i as well as *aggregate reputation* $Q_t \equiv G^{-1}(\int_{\mathcal{I}} G(q_t^i) di)$, an index of overall reputation in the economy. The time-invariant function f thus summarizes the economy-wide "demand system," relating an expert's inflow of problems to the reputation profile of all experts. The function $G(\cdot)$ aggregates the individual reputations. We impose mild assumptions on f and G :

Assumption 1 (Regularity)

- (i) $f(q, Q)$ is continuously differentiable in both arguments, $f_1(q, Q) > 0$, $f_2(q, Q) < 0$, $f_1(q, q) + f_2(q, q) \geq 0$.
- (ii) $h(q, Q) \equiv \frac{f_1(q, Q)}{f(q, Q)}$ decreases in q , and increases in Q but not too fast.
- (iii) $G(\cdot)$ is continuously differentiable, and $G'(\cdot)$ is strongly positive.

Condition (i) puts smoothness and monotonicity requirements on f . The demand for expert i 's service increases in their reputation but decreases in aggregate reputation. Moreover, if the reputation of all experts increases, expert i 's demand does not decrease. Condition (ii) imposes restrictions on the growth rate of expert i 's demand in the expert's reputation q , reflecting the sensitivity of demand f to q . It

¹⁰Note we implicitly assume that the speed of decay is independent of the outcome of the problems.

requires this growth rate to decrease in the expert's reputation q so demand cannot grow more than exponentially. Finally, the higher the aggregate reputation Q , the more sensitive expert i 's demand f 's is to their own reputation q . Condition (iii) guarantees that the aggregate reputation is meaningful and regular.

The combination of G and f allows for a general set of demand systems including some common cases. For example, if $G(\cdot)$ is the identity function and f is linear in q and Q , a linear demand system results. $G(q) = q^\sigma$ ($\sigma > 1$) and $f(q, Q) = q^\sigma Q^{1-\sigma}$ give rise to a CES-type system. When $G(q) = e^{\sigma q}$ ($\sigma > 0$) and $f(q, Q) = e^{\sigma(q-Q)}$, f and Q describe a Logit-type system. All these systems satisfy [Assumption 1](#).

Because each expert's weight in the economy is "small," each expert treats the law of motion of Q_t as given, but affects q_t^i directly through their choice of the referral rate r_t^i . The record \mathcal{R}_t^i thus evolves according to the law of motion:

$$\frac{dm_t^i}{dt} = \phi \cdot (f(q_t^i, Q_t) - m_t^i), \quad (1)$$

$$\frac{dq_t^i}{dt} = \phi \cdot \frac{f(q_t^i, Q_t)}{m_t^i} \cdot (\eta_t^i - q_t^i). \quad (2)$$

Payoffs. For each expert, we normalize the earnings for treatment and diagnosis as $\pi \in (0, 1)$ and $1 - \pi$ respectively, and—to ease exposition—refer to them as fees. We assume these fees reflect net revenues accounting for various costs of the experts. If an expert refers a problem to another one, the referring expert earns only the diagnostic fee, while the expert being referred to receives only the treatment fee.

Let φ_t^i be the total number of referrals expert i receives in period t .¹¹ Then, expert i of type j derives flow payoff

$$f(q_t^i, Q_t) \cdot (1 - p_{-j} \pi r_t^i) + \pi \varphi_t^i,$$

where the first term is the profit from diagnosing and treating problems they receive in period t minus the treatment fees for problems referred to other experts. The second term describes treatment fees from problems received via referrals.

¹¹Recall that referred problems always fall into the recipient's category, as referring a problem to an expert of the same type causes a loss of the treatment fee without a corresponding reputation gain.

All experts discount the future at the same rate $\rho > 0$. Therefore, the discounted future payoff of type j expert i amounts to

$$V_0^i = \int_0^\infty f(q_t^i, Q_t) \cdot (1 - p_{-j}\pi r_t^i) e^{-\rho t} dt + \int_0^\infty \pi \varphi_t^i e^{-\rho t} dt.$$

The entire history of records up to time t is common knowledge among experts.

Experts choose their referral rate at time t in order to maximize their future discounted payoffs. The dilemma each expert faces is the following. When referring a mismatched problem, they forfeit the treatment fee π . On the other hand, referring a problem to the right type of expert increases the probability of a successful outcome, which in turn raises the referring expert's reputation for solving problems successfully, thereby generating a higher demand for their services.

Strategies & Equilibrium. We examine referrals as strategic instruments to trade off present and future income mediated through reputation. In particular, it is not quid-pro-quo deals or tit-for-tat strategies among acquainted experts we are interested in. We therefore focus on Markov strategies throughout. A Markov strategy of expert $i \in \mathcal{I}_j$ is a time-invariant function $\sigma^i : L^1([0, 1], [0, 1]) \times L^1([0, 1], [0, 1]) \mapsto [0, 1]$, mapping the profile of records $\{(m_t^i, q_t^i)\}_{i \in \mathcal{I}}$ into their time t referral rate r_t^i .

A Markov equilibrium of the game is a strategy profile $\{\sigma^i\}_{i \in \mathcal{I}}$ such that given any starting profile of records $\{(m_0^i, q_0^i)\}_{i \in \mathcal{I}}$, strategy σ^i is a best response for expert i given the strategies $\{\sigma^{i'}\}_{i' \neq i}$ of all other experts, for all $i \in \mathcal{I}$. We specifically focus on *steady states* of the game, i.e., a tuple $\{(m_*^i, q_*^i, r_*^i)\}_{i \in \mathcal{I}}$ such that evaluated there, for all $i \in \mathcal{I}$, we have $\frac{dm_t^i}{dt} = 0$, $\frac{dq_t^i}{dt} = 0$, and $r_*^i = \sigma^i(\{(m_*^{i'}, q_*^{i'})\}_{i' \in \mathcal{I}})$.

Finally, as referrals add to the instantaneous payoff of the receiving doctor but do not affect their reputation, the identity of the referrals' recipients is immaterial for the expert's optimization problem and thus equilibrium behavior. We focus on outcomes that are symmetric among experts of the same type. As such, we assume that each expert of type j is equally likely to receive a problem referred by an expert of

type $-j$.¹² As a result, type j expert i 's future discount payoff amounts to

$$V_0^i = \int_0^\infty f(q_t^i, Q_t) \cdot (1 - p_{-j}\pi r_t^i) e^{-\rho t} dt + \frac{1}{w_j} \int_0^\infty \int_{\mathcal{I}_{-j}} f(q_t^{i'}, Q_t) p_j \pi r_t^{i'} e^{-\rho t} di' dt.$$

3 Equilibrium Analysis

In this section, we use recursion to analyze Markov equilibria of the referral game described above, and then establish the existence of a unique steady state.

3.1 Recursive Method

The game is complicated as it involves a continuum of players and state variables. However, to identify steady state equilibria, we fix the records of all other experts $i' \neq i$ at some conjectured steady state values and study the behavior of a generic expert i . If it is optimal for expert i to choose a referral rate r^i that induces a constant record, then a partial steady state (i.e., for the expert) is found.

The Problem. Suppose $(m_t^i, q_t^i, r_t^i) = (m_*^i, q_*^i, r_*^i), \forall i' \neq i, t \geq 0$, and define $Q_* \equiv G^{-1}(\int_{\mathcal{I}} G(q_*^i) di)$. Now that $Q_t = Q_*$ is the constant aggregate reputation of experts in the economy, expert i of type j faces a recursive decision problem:

$$\begin{aligned} V^i(m^i, q^i) &= \max_{r^i \in [0,1]} f(q^i, Q_*) \cdot (1 - p_{-j}\pi r^i) + \frac{1}{w_j} \int_{\mathcal{I}_{-j}} f(q_*^{i'}, Q_*) p_j \pi r_*^{i'} di' \\ &+ \mu \cdot \frac{f(q^i, Q_*)}{m^i} \left(r^i \bar{q}_j + (1 - r^i) \underline{q}_j - q^i \right) \frac{\partial V^i(m^i, q^i)}{\partial q^i} \\ &+ \mu \cdot (f(q^i, Q_*) - m^i) \frac{\partial V^i(m^i, q^i)}{\partial m^i}, \end{aligned} \quad (3)$$

where $\bar{q}_j \equiv p_j \eta_j^j + p_{-j} \eta_{-j}^{-j}$, $\underline{q}_j \equiv p_j \eta_j^j + p_{-j} \eta_j^{-j}$, and $\mu \equiv \frac{\phi}{\rho} > 0$ is the decay of records relative to discounting. There are four terms on the right hand side of this Hamilton-Jacobian-Bellman (HJB) equation. The first two are flow payoffs from expert i 's retained demand and their received referrals. The third term is the drift of the value

¹²Note that this symmetry assumption is innocuous for equilibrium referral rates, experts' records, and overall demand in the economy and only additively affects individual expert payoffs.

function due to changes in reputation q^i , and the last is the drift due to changes in their record's thickness m^i . Since expert i 's referral rate r^i lies between 0 and 1, we know $r^i\bar{q}_j + (1 - r^i)\underline{q}_j \in [\underline{q}_j, \bar{q}_j]$. Therefore, if $q_0^i \in [\underline{q}_j, \bar{q}_j]$, then $q_t^i \in [\underline{q}_j, \bar{q}_j]$ for all $t \geq 0$. Likewise, since $f(\cdot, Q_*)$ is strictly increasing, if $m_0^i \in [f(\underline{q}_j), f(\bar{q}_j)]$, then $m_t^i \in [f(\underline{q}_j), f(\bar{q}_j)]$ for all $t \geq 0$. As a result, expert i 's record is naturally bounded in both dimensions, i.e., in its thickness m^i and reputation q^i .

The maximization in (3) with respect to r^i requires that

$$r^i \begin{cases} = 1 & \text{if } \frac{\mu}{m^i} \frac{\partial V^i(m^i, q^i)}{\partial q^i} > \frac{p_{-j}\pi}{\bar{q}_j - \underline{q}_j}, \\ \in [0, 1] & \text{if } \frac{\mu}{m^i} \frac{\partial V^i(m^i, q^i)}{\partial q^i} = \frac{p_{-j}\pi}{\bar{q}_j - \underline{q}_j}, \\ = 0 & \text{if } \frac{\mu}{m^i} \frac{\partial V^i(m^i, q^i)}{\partial q^i} < \frac{p_{-j}\pi}{\bar{q}_j - \underline{q}_j}. \end{cases} \quad (4)$$

Note that although the record's thickness m^i is a 'latent' state in the sense that it does not directly affect demand, it is crucial for the incentive to refer indirectly affecting future demand through reputation. The reason is that m^i reflects the record's inertia, i.e., how sensitive the expert's reputation q^i is to current referrals. When m^i is small, the record is thin, so that a current referral boosts q^i significantly. On the contrary, when m^i is large, the record is too thick to substantially reflect current referrals. That is to say, a larger m^i dampens the incentive to refer a mismatched problem.

Interpretation. A heuristic interpretation of the dynamics of this two-dimensional system is as follows. Assume an expert refers all mismatched problems, thereby maximally increasing their reputation q^i given their record's thickness m^i . This, in turn, increases the expert's demand. As a result, the expert's record's thickness increases to a point at which the expert is no longer willing to refer. At this point their reputation q^i starts to decline, followed by a subsequent decrease in thickness m^i . This negative feedback loop suggests the existence of an equilibrium.

Existence. Formally, the existence of a unique weak solution (or viscosity solution) to the problem given by Equation (3) is guaranteed by [Crandall, Ishii, and Lions \(1992\)](#), where the HJB equation is satisfied everywhere except on a zero-measure set. Given the unique solution, the optimal referral policy r^i defined by Equation

(4) is an upper hemi-continuous correspondence. This guarantees the existence of a fixed point (m_*^i, q_*^i, r_*^i) such that when evaluated there, $\frac{dm_*^i}{dt} = 0$, $\frac{dq_*^i}{dt} = 0$, and r_*^i satisfies Equation (4). By definition, this is a steady state for expert i in the partial equilibrium when the aggregate reputation Q_* is fixed.

Lemma 1 (Existence)

There exists a unique weak solution $V^i(\cdot, \cdot)$ to (3). Moreover, there exists a steady state (m_^i, q_*^i, r_*^i) of the partial differential equations defined by the weak solution and the associated policy correspondence $r^i(\cdot, \cdot)$.*

The solution of the HJB equation is sometimes ‘weak’ because the value function V^i may not be differentiable on zero measure sets, causing r^i to be undefined. However, extending the definition of a derivative by locally perturbing state variables and convexifying the set of resulting derivatives ensures that the policy correspondence r^i is again well defined. A Kakutani-style argument then implies the existence of a fixed point, i.e., a steady state. We deal with uniqueness below in [Proposition 1](#).

3.2 Steady State

Partial Equilibrium. For a given aggregate reputation in the economy Q_* , [Lemma 1](#) establishes that expert i ’s decision problem admits a steady state. We then establish uniqueness in [Proposition 1](#) below, and present conditions under which there is no, partial or full referral by expert i in the unique steady state.

Proposition 1 (Steady State for an Expert)

Given Q_ , there exists a unique steady state (m_*^i, q_*^i, r_*^i) for expert i of type j , where:*

$$m_*^i = f(q_*^i, Q_*), \tag{5}$$

$$q_*^i = r_*^i \bar{q}_j + (1 - r_*^i) \underline{q}_j, \tag{6}$$

$$(1 - p_{-j} \pi r_*^i) h(q_*^i, Q_*) \frac{\bar{q}_j - \underline{q}_j}{p_{-j}} \frac{\mu}{1 + \mu} - \pi \begin{cases} > 0 & \text{if } r_*^i = 1, \\ = 0 & \text{if } r_*^i \in [0, 1], \\ < 0 & \text{if } r_*^i = 0. \end{cases} \tag{7}$$

Intuitively, the condition that determines expert i 's steady state referral rate r_*^i reflects a dynamic trade-off between today's and tomorrow's payoffs. If the expert refers more aggressively, i.e., refers more mismatched problems to experts of the other type in order to increase their reputation, then the first term in Equation (7) reflects their marginal benefit from referral, that is, the expected profit expert i makes for one problem they initially receive, $1 - p_{-j}\pi r_*^i$, multiplied by the relative growth in their demand $h(q_*^i, Q_*) \frac{\bar{q}_j - q_j}{p_{-j}}$ resulting from increased reputation. Note that $\frac{\bar{q}_j - q_j}{p_{-j}} = \eta_{-j}^{-j} - \eta_j^{-j}$ reflects the gains in the current success rate per referral. This marginal benefit is then normalized by a factor $\frac{\mu}{1+\mu} = \frac{\phi}{\rho+\phi}$. This is because a higher ϕ translates into higher reputation gains per additional referral. On the other hand, a higher value $\rho + \phi$ reflects a decline in the future reward for higher reputation as it is both discounted more and decays faster. The second term represents the marginal costs of an additional referral, which is simply the lost treatment income π . The expert's steady state referral rate is interior only if the expert is indifferent between the marginal gains and costs of an additional referral.

Figure 1 illustrates the feedback loop that gives rise to the existence of a partial steady state solution. The solid curve is the exact locus of (m, q) such that the new inflow of problems offsets the decay in thickness. The dashed curve is the switching surface where the expert discretely shifts from referring all mismatched problems to referring none. As long as the thickness of expert i 's record is below the dashed curve, they find reputation sufficiently responsive to current referrals and thus prefer referring all mismatched problems. This, in turn, increases their future reputation. As a result, instantaneous demand increases and eventually so does the thickness of expert i 's record. As such, the payoff from referring mismatched problems diminishes, eventually to a point at which the expert ceases to refer.

General Equilibrium. Naturally, all experts of the same type j have the same steady state. Let (q_{j*}, m_{j*}, r_{j*}) denote the steady state of any expert of type $j = A, B$. We now invoke our insights from the partial equilibrium in **Proposition 1** and consistency of the aggregate reputation Q_* with individual expert behavior, i.e.,

$$G(Q_*) = w_A G(q_{A*}) + w_B G(q_{B*}), \quad (8)$$

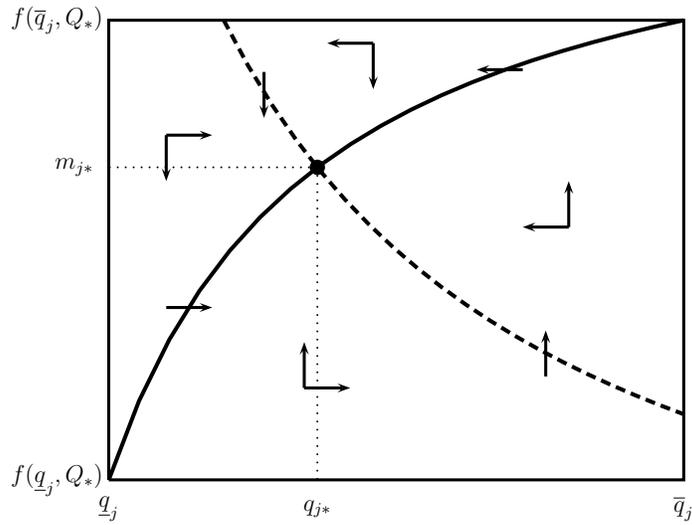


Figure 1: Phase diagram of the (m, q) system. The solid curve represents $m = f(q, Q_*)$, and the dashed curve is the switching surface from $r = 1$ (below) to $r = 0$ (above). The steady state is given by the intersection of the two curves.

to establish a unique steady state in the general equilibrium of the economy.

Proposition 2 (Steady State in the Economy)

The economy admits a unique steady state $\{(m_{j}, q_{j*}, r_{j*})\}_{j=A,B}$, characterized by (5)-(8).*

The existence and uniqueness of a general equilibrium steady state is established by invoking the Contraction Mapping Theorem. Given any level of aggregate reputation Q^* , we can derive q_{A*} and q_{B*} from the partial equilibrium, and then update Q^* according to Equation (8). Under [Assumption 1](#), such a mapping is a contraction.

Welfare. Note that a steady state of Markov equilibria admits a straightforward interpretation of the variables of our model as measures of overall welfare. First consider the experts' respective reputation q_{j*} and q_{-j*} . By definition of a steady state, the referral rate of any expert is constant over time. As a result, q_{j*} (resp. q_{-j*}) indicates the fraction of problems (initially received) that a type j ($-j$) expert solves. Note, that—holding everything else fixed—for a given referral rate of type $-j$ experts r_{-j*} , a higher referral rate by type j experts unambiguously increases q_{j*} , and thereby their overall success rate in solving problems. In fact, the first best outcome requires both types of experts to refer all mismatched problems.

It follows, however, from [Proposition 1](#) and [Proposition 2](#) above that in the economy's unique steady state, experts generally do not refer all mismatched problems they receive. We emphasize this inefficiency in [Corollary 1](#) below.

Corollary 1 (Inefficient Referrals)

In the unique steady state of the economy, experts in general do not refer problems efficiently.

In the next section we investigate the factors that push the economy towards (or away from) efficient referral behavior.

4 Comparative Statics

In this section, we analyze the primitives of the economy and how they affect equilibrium referral behavior by experts and their steady state reputation. This allows us to better understand which industry characteristics pose challenges to the market in efficiently allocating problems across experts. First, we consider the effects of experts' respective abilities in treating problems, and discuss the potential effects of experts adopting artificial intelligence tools. Thereafter we analyze how the properties of the experts' record and its evolution set the stage for the experts' referral game. Finally, we investigate how the fee structure for experts' services affects outcomes. When describing actions and outcomes throughout this section, we refer to steady state values. Finally, to aid readability we write j and $-j$ to refer to all experts of their respective types.

4.1 Expert Abilities

Leading Ability. Expert abilities are at the core of our model. They are synonymous with direct success rates when experts treat a specific type of problem, but also affect experts' eventual overall success rates in solving problems through their effect on referral rates. We start out by analyzing the effects of an increasing type j expert's ability in treating problems of their category η_j^j .

Proposition 3 (Leading Ability)

Consider type j experts' ability in treating their problems $\eta_j^j \in (\eta_{-j}^j, 1]$. All experts' steady state reputation, i.e., q_{j*} and q_{-j*} , increase in η_j^j .

Proposition 3 establishes that a higher ability of expert j in treating problems of their category η_j^j increases the reputation of both types of experts, i.e., q_{j*} and q_{-j*} . There are both direct and indirect channels through which η_j^j affects reputation. The direct effect is straightforward. Type j is better in treating their own type of problems, and type $-j$ benefits when referring mismatched problems. The indirect effect via equilibrium referral rates, however, is more subtle. On the one hand, increased reputation of both types of experts increases aggregate reputation Q_* , which in turn provides further incentives to refer. On the other hand, there is a countervailing force. An increased reputation leads to a thicker record. As a result, an expert will find reputation growth marginally less rewarding. These forces combine to an ambiguous indirect effect, and the referral rates of both expert types may either increase or decrease in response to a higher η_j^j . Nevertheless, the direct effect is strong enough to guarantee an increase in reputations, and thus in the overall success rate in solving problems in the economy.

Lagging Ability. Next we investigate the equilibrium effects of an increasing type j expert's ability in treating mismatched problems η_j^{-j} .

Proposition 4 (Lagging Ability)

Consider type j experts' ability in treating mismatched problems $\eta_j^{-j} \in [0, \eta_{-j}^{-j})$.

(i) Type j experts' steady state referral rate r_{j*} decreases in η_j^{-j} . Their steady state reputation q_{j*} achieves its maximum when expert type j always refers, decreases in η_j^{-j} when type j experts partially refer, and increases in η_j^{-j} when they never refer.

(ii) Type $-j$ experts' steady state referral rate r_{-j*} and reputation q_{-j*} achieve their respective maximum when expert type j always refers, decreases in η_j^{-j} when type j experts partially refer, and increases in η_j^{-j} when they never refer.

Note a marginal change in η_j^{-j} has no effect on type j 's referral behavior if their referral rate is given by a corner solution to begin with. When $r_{j*} = 1$, a change in

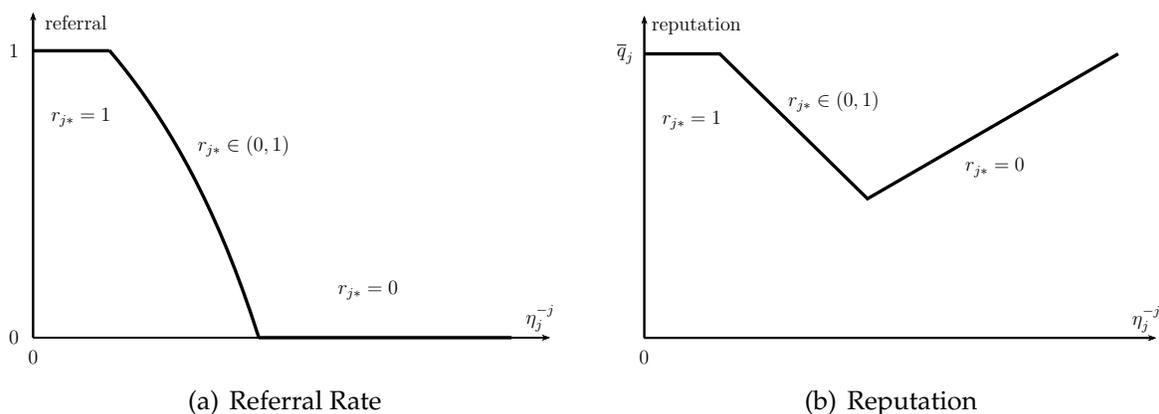


Figure 2: Type j 's steady state referral rate r_{j*} and reputation q_{j*} as a function of their lagging ability η_j^{-j} .

η_j^{-j} has no bearing on equilibrium reputation q_{j*} either. When $r_{j*} = 0$, on the other hand, η_j^{-j} mechanically raises type j 's equilibrium reputation. If their referral rate is interior, however, as stated in [Proposition 4\(i\)](#), an increase in their ability to treat mismatched problems reduces an expert's incentive to refer problems to experts of the other type. Furthermore, this indirect effect dominates the direct effect from an increased ability to treat mismatched problems. Part (ii) notes the impact of η_j^{-j} on the other type $-j$'s referral rate and reputation. This effect is purely mediated through the general equilibrium channel: $-j$'s incentive to maintain their reputation increases with the aggregate reputation Q_* , which in turn depends on q_{j*} . As a result, $-j$'s referral rate and reputation move in the same direction as j 's reputation.

[Figure 2](#) illustrates the overall effects of an increase in type j 's lagging ability on their referral rate and reputation. Panel (a) visualizes the response of type j 's referral rate. As discussed above, there is no effect if a type j expert never or always refers in the first place. Otherwise, their referral rate decreases as the expert becomes better at treating mismatched problems. Panel (b) graphically represents the effect of an increased lagging ability on an expert's reputation. If the expert always refers to begin with, there is no effect, while their reputation mechanically increases if they refer no mismatched problem. However, in an interior solution, where the expert refers some but not all mismatched problems, their steady state success rate (reputation) decreases.

In summary, suppose the steady state of the Markov equilibrium is not a corner solution, i.e., j 's referral rate r_{j^*} in equilibrium is interior, meaning type j experts refer some but not all mismatched problems they receive. Then, an increase in j 's ability to treat mismatched problems lowers all experts' referral rates as well as their respective reputation, i.e., their discounted success rates in solving problems.

Artificial Intelligence. An interesting application of [Proposition 4](#) relates to the advent of Artificial Intelligence (AI) in the workplace. Contemporary research (see, e.g., [Brynjolfsson, Li, and Raymond, 2023](#)) strongly suggests that the adoption of AI tools substantially increases the output of less productive workers but only minimally affects experienced or high-ability ones. Importantly, this distinction remains true for college-educated professionals ([Noy and Zhang, 2023](#)) and, more generally, for high skilled work ([Cui, Demirer, Jaffe, Musloff, Peng, and Salz, 2024](#)). As such, when interpreting the impact of adopting AI in the workplace as an increase of experts' lagging ability, we have the following result.

Corollary 2 (Artificial Intelligence)

Suppose type j experts adopt AI tools that increase η_j^{-j} . If type j refers some but not all mismatched problems to type $-j$ experts before and after the adoption, AI unambiguously lowers the steady state referral rates and the reputation of all experts.

[Corollary 2](#) suggests that AI may impact productivity in a way that—to the best of our knowledge—has so far been overlooked in the popular and academic discourse. If the access to AI tools fosters mainly secondary rather than professionals' primary skills, it may compel these professionals to take on more problems that are not a good match for their expertise, ultimately increasing the rate of failure. Also note that, due to the uniqueness of steady state, [Corollary 2](#) can be applied sequentially to analyze the effects of both expert types honing their lagging ability. In other words, if both η_j^{-j} and η_{-j}^j increase, both changes affect referral rates, and thus the economy's overall success rates in solving problems negatively, as long as the equilibrium referral rates remain interior.

Specialization. [Proposition 3](#) and [Proposition 4](#) establish the welfare consequences of experts honing their leading and lagging ability. It follows that a hypothetical so-

cial planner—interested in increasing the economy’s overall success rates in solving problems (and potentially increasing overall demand)—prefers experts to further develop their respective specialization rather than increasing their success rate in treating mismatched problems. As we show below, however, this is not necessarily the case for the experts themselves. In fact, under fairly general conditions, experts may be compelled to invest in their lagging rather than their leading ability.

Proposition 5 (Specialization)

Suppose demand is not too elastic in aggregate, i.e., $|f_1(q, Q) + f_2(q, Q)|_{Q=q}$ and $|h_1(q, Q) + h_2(q, Q)|_{Q=q}$ are sufficiently small. Then, if the mass w_j of type j experts is sufficiently large, these experts marginally benefit more from an increase of η_j^{-j} than η_j^j .

Proposition 5 states that when the overall demand for experts’ services is not too elastic, and there is sufficiently many experts of one type in the economy, these experts are more interested in honing their lagging rather than their leading ability. Intuitively, when the aggregate elasticity is small, then a universal increase or decrease of reputation does not significantly affect individual demand. This, combined with the fact that there is sufficiently many experts of one type, leads to a scenario where reputation cannot discipline this type’s behavior because aggregate reputation is sufficiently similar to theirs. When η_j^j increases, type $-j$ refers more, but their sheer volume is insufficient. When η_j^{-j} increases, the group j refers less often and earns more treatment profit, which is the decisive factor discouraging referrals.

While we do not explicitly model costly investment into experts’ abilities, it is straightforward that a well-behaved cost function induces these experts to narrow the gap between their abilities rather than invest in their specialty. This is an alarming result as it pertains to a situation in which (i) there are no viable substitutes for the experts’ services, and (ii) the expert type that chooses investment that is detrimental to overall welfare is pervasive.

4.2 Reputation

In this section we examine the equilibrium effects when changing how the experts’ record disciplines their referral behavior. First, consider $\mu \equiv \frac{\phi}{\rho}$, the decay rate of

records relative to the discount rate of time.¹³ It measures how sensitive the record is to today's actions vs. the importance of the record itself. Second, consider the growth rate $h(\cdot, Q)$ of the demand for the experts' services with respect to aggregate reputation. It reflects how the market reacts to changes in the expert's reputation.

Proposition 6 (Reputation)

Consider μ and $h(\cdot, Q)$. All experts' steady state referral rates and reputation increase in μ . They also increase if $h(\cdot, Q)$ increases everywhere for any fixed Q .

The results in Proposition 6 are intuitive. When the expert's record decays faster (larger ϕ), i.e., when it becomes more sensitive to current performance, experts refer more often to increase their reputation. Equally, if the value of tomorrow increases (smaller ρ), so do experts' referral rates, and therefore their reputation. Likewise, if an increase in today's reputation increases tomorrow's demand (higher $h(\cdot, Q)$) more significantly, referral rates and reputation increase.

4.3 Fee Structure

We now focus on the fee structure of the expert economy, that is, the experts' fees for diagnosing $(1 - \pi)$ vs. treating (π) a problem. A higher treatment fee π affects an expert's incentives to refer problems in different ways. Directly, it discourages the expert from referring a problem due to the heightened revenue loss. Indirectly, a higher π corresponds to a lower diagnostic fee, implying lower returns to reputation. An expert is thus further discouraged from referring.

Proposition 7 (Fee Structure)

Fix all parameters except π . A higher treatment fee π lowers all experts' steady state referral rates and reputation.

This result conveys bad news from a welfare perspective. While we do not explicitly model the experts' initial investment in their abilities such as education, it appears

¹³Note that μ strictly increases in ϕ but strictly decreases in ρ . As a result, the qualitative comparative statics below are fully informative for the respective parameters.

reasonable to assume that the treatment fee serves as an important motivation for experts to develop the necessary skills to enter the field. [Proposition 7](#), however, suggests that the higher the treatment fee, the lower the success rates of both types—and thus the overall economy’s—in solving problems.

5 Referral Alliances and Partnerships

Market power. Horizontal mergers, i.e., mergers between firms that produce substitutes or complements but operate on the same level of the value chain, often increase the market power of the merged entity, allowing it to increase its prices for a given level of product quality, thereby harming social welfare. This is no different for mergers or partnerships among skilled professionals. [Dunn and Shapiro \(2014\)](#) for example analyze market for physicians’ services and find that market concentration increases physician fees. In fact, mergers in the medical field may even lead to higher market prices when the merging entities are not substitutes but scale increases their bargaining power vis-à-vis insurers ([Dafny, Ho, and Lee, 2019](#)).

Contracts. Below, however, we show that if a market is subject to misallocation due to a lack of referrals, *referral alliances* and *partnerships* based on mutual contracts between experts may contribute to alleviate the problem. A contract between experts specifies a pair of treatment revenue shares (y_a, y_B) , and a pair of lump-sum transfers (T_A, T_B) . To be specific, $y_j \in [0, 1]$ is the share of treatment income π an expert is entitled to when referring a problem of category j , with the recipient earning $(1 - y_j)\pi$ when treating the problem. In other words, $y_j\pi$ is the price for a referral of a category j problem. $T_j \in \mathbb{R}$ is a referral-independent redistributive payment from each type j expert. Budget balance requires $n_A T_A + n_B T_B = 0$, where n_j is the number of type j experts entering the contract. We refer to experts entering any such contract as a referral alliance and to those specifically entering a contract with referral-independent revenue sharing, i.e., $|T_j| > 0$ for $j \in \{A, B\}$, as a partnership. We assume throughout that entering a referral alliance per se does not affect an expert’s ability, the demand for their services, nor the law of motion of their records. Below we show that higher prices for referrals naturally encourage

increased referral activity.

Lemma 2 (Referral Prices)

Suppose that a set of experts, containing at least one expert of each type, has entered a referral alliance. Then—holding the behavior of all experts outside the alliance constant—all type j experts within the alliance weakly increase their steady state referral rates in response to a higher y_{-j} . Moreover, the referral alliance guarantees efficient referral behavior, i.e., $r_{j*} = r_{-j*} = 1$ if and only if $y_j \geq \underline{y}_j$ for $j = A, B$, where:

$$\begin{aligned} \underline{y}_j &\equiv 1 - \frac{\mu(\bar{q}_{-j} - \underline{q}_{-j})h(\bar{q}_{-j}, \bar{Q})}{p_{-j}\pi \left(1 + \mu + \mu(\bar{q}_{-j} - \underline{q}_{-j})h(\bar{q}_{-j}, \bar{Q})\right)} < 1, \\ \bar{Q} &\equiv G^{-1}(w_A G(\bar{q}_A) + w_B G(\bar{q}_B)). \end{aligned} \tag{9}$$

The intuition behind [Lemma 2](#) is as follows. Without entering a referral alliance, a type j expert’s reward from referring mismatched problems stems solely from the positive demand effect mediated through an increase in their reputation. When entering a contract with other experts, however, the expert enjoys the same benefit from referring problems but, additionally, earns y_{-j} per referred problem. As a result, the expert exclusively refers to experts within the alliance, and the incentive to refer increases in the referral price y_{-j} . If both y_j and y_{-j} are sufficiently large, all experts within the alliance refer all mismatched problems they receive.

Forming an Alliance. In order for experts to establish a referral alliance, however, jointly entering such a contract needs to be a profitable deviation from the initial steady state for each expert. To this end, suppose no expert enters into a referral alliance but a finite set \mathcal{I}_P that consists of at least one expert of each type. Also, for simplicity, let type j experts within the alliance refer mismatched problems to each type $-j$ expert with equal probability. Then, mutually joining a referral alliance is desirable for every expert in \mathcal{I}_P even if $T_A = T_B = y_A = y_B = 0$.

Proposition 8 (Forming a Referral Alliance)

Suppose $n_A \geq 1$ type- A experts and $n_B \geq 1$ type- B experts have entered a referral alliance

with $T_A = T_B = 0$. Then—for any behavior of experts outside the alliance—all experts within benefit from the alliance regardless of y_A and y_B .

An expert prefers entering into a referral alliance for two reasons. First, due to treatment revenue sharing of referred problems, the merger confines referrals exclusively within the alliance. This increases the number of referrals an expert within receives, growing their payoff. Second, treatment revenue sharing increases an expert's payoff from referring a mismatched problem increasing their referral rate, and thereby their reputation. While an increased referral rate goes hand in hand with revenue loss, the expert could revert to the steady state referral rate of her type outside the partnership still benefiting from an increased number of referrals they receive. As a result, the expert's overall benefit from joining a referral alliance must be positive.

Pervasive Partnerships. Understanding the benefits of a joint deviation into a referral alliance allows us to examine the economy-wide effects of competing alliances. It turns out that there is a steady state of a Markov equilibrium in which all experts join partnerships, referrals are efficient economy-wide, all experts are better off than in the equilibrium in which individual experts compete and overall demand is maximized. Moreover, no single partnership holds market power in the economy.

Proposition 9 (Pervasive Partnerships)

Suppose every expert in the economy has entered a partnership, and every partnership consists of finite number of experts featuring both types. Then, for every partnership, there exists a Pareto-efficient partnership contract that guarantees efficient referral behavior, and the partnerships are stable against joint deviations by any finite number of experts.

[Proposition 9](#) suggests that partnerships that are small in size are incentive-compatible for experts and restore efficiency in the economy. Similar to the proposition statement, one can show there also exists an equilibrium in which every expert has entered a referral alliance in which $T_A = T_B = 0$. While it follows from [Proposition 8](#) that every experts wants to enter an alliance, some experts may be worse off in an economy with pervasive referral alliances than in an economy in which individual experts compete. As such, these experts may find themselves in a Prisoner's

Dilemma type situation. This is where redistributive transfers come into play that distinguish a partnership from a simple referral alliance. There always exists an equilibrium with pervasive partnerships in which experts that overly benefit from the partnership structure transfer just enough of their income to those experts who are harmed relative to individual competition. In such an equilibrium, experts refer all their mismatched problems, and overall demand and the success rate in solving problems in the economy are maximized.

This is an important point that should be considered when regulating markets with specialized professionals and uninformed clients. In fact, our results suggest that referral alliances and partnerships of limited size may be essential in fostering efficient referrals among experts in the economy. As long as these alliances are small in size relative to the overall market, welfare loss through market power should be negligible. Moreover, partnerships that feature sharing of revenue beyond referrals can ensure that all experts are strictly better off than in an economy of individual competition with inefficient referral behavior.

Finally, note that a Pareto-efficient partnership contract necessarily maximizes the partnership's total revenue given the structure of the economy. As a result, partnerships do not only hold the potential to resolve a lack of referrals but may also contribute to alleviate problems related to specialization and the fee structure discussed in Section 4. The partnership's value increases in their specialists' abilities and thus may encourage experts to invest in their leading rather than in their lagging ability. Moreover, due to revenue sharing, efficient referral behavior does not necessitate low treatment fees when experts have entered partnership contracts.

6 Discussion

The idea that firms incur losses in the short-run building up reputation to increase the future demand for their products dates back to [Shapiro \(1983\)](#). This is no different in markets for professional services. Medical providers face incentives not to treat cases that are not within their specialty, as those known for better performance in the past attract more patients ([Chandra, Finkelstein, Sacarny, and Syversen, 2016](#)).

Referrals in the healthcare setting also stem from a fear of malpractice suits, leading to reputation loss (Franks, Williams, Zwanziger, Mooney, and Sorbero, 2000). Moreover, referrals impact both the costs and outcomes of health care significantly (Eftekhari, Yaraghi, Gopal, and Ramesh, 2023). Likewise, referrals are common among competing lawyers (Liu and Hyman, 2024), and many law firms lay out in detail why, when and how to refer cases to other lawyers.¹⁴

As our experts are different types of specialists, one may argue that our model does not reflect medical services well relying on gatekeepers such as family physicians. In fact, while gatekeeping is uncommon in some of the world’s biggest economies (Gonzalez, 2010), and referrals between specialists are common not only in the U.S. (An, O’Malley, and Rockmore, 2018), we argue that our model also accommodates referrals between general practitioners (GPs) and specialists. This is because GPs can be seen as specialists in common diseases they can treat more effectively because of their experience. In fact, An, O’Malley, and Rockmore (2018) shows that specialists refer plenty of patients to family practices.¹⁵

Consistent with the main trade off in our model, there is strong evidence that increased income from treatment directly lowers referral rates (Barnett, Song, and Landon, 2012; Ringberg, Fleten, Deraas, Hasvold, and Førde, 2013; Godager, Iversen, and Ma, 2015). Furthermore, referrals increase in specialization and specialist skill (Barnett, Song, and Landon, 2012; Ringberg, Fleten, Deraas, Hasvold, and Førde, 2013). Iversen and Ma (2011) and Godager, Iversen, and Ma (2015) argue that competition among medical providers increases referral rates. Similarly, in our model when the reputation effect on demand increases, i.e., when it more important to differentiate oneself, such as under heightened competition, referral rates increase.

Drawing on steady states of Markov equilibria to analyze the effects of market con-

¹⁴See <https://mccreadylaw.com/blog/referring-cases-lawyers/> as a prime example urging lawyers not to “... accept a case in an unfamiliar area ...” because “At best, it will make [them] look bad, at worst, it may end up being a malpractice claim.”

¹⁵On a different note about gatekeepers, in our model, even if type j has an absolute advantage in treating either problem, it is not guaranteed that a type $-j$ expert always refers turning into a pure gatekeeper (i.e., a diagnostician). Just as in the analysis in Section 3, an expert may not be willing to give up treatment fees for better reputation, even they lag in both abilities. In other words, for experts to become pure gatekeepers in equilibrium, they need to be substantially worse than experts of the other type in solving both problems of both categories.

ditions is consistent with findings of [Franks, Zwanziger, Mooney, and Sorbero \(1999\)](#) and [Franks, Williams, Zwanziger, Mooney, and Sorbero \(2000\)](#) showing that health providers' referral rates are fairly stable over a provider's life cycle. Recently, however, the number of referrals in the U.S. healthcare sector, including referrals from specialists to specialists, sharply increased ([Barnett, Song, and Landon, 2012](#)). This is consistent with increased specialization in the market, but also with an increase of experts' reputational concerns in the age of readily available online reviews.

Finally, there is vast evidence of partnerships and referral alliances in markets for professional services. [Nam, Gruca, and Tracy \(2010\)](#) show that not only referrals are common in law, insurance, investment banking and consulting, but so are referral alliances and networks. [Gonzalez and Rizzo \(1991\)](#) detail that medical doctors in group practices commonly refer to each other, while [Garicano and Hubbard \(2009\)](#) argue that law firms started hiring from more diverse fields of the law to facilitate within-practice referrals. More generally, in a study of business professionals referring professionals to financial institutions, [Business Development Bank of Canada \(2013\)](#) find that experts are more likely to refer when referrals are reciprocal.

7 Conclusion

Experts increase demand for their services by cultivating a higher reputation for being an able problem solver. One way to strengthen their reputation is for an expert to refer problems not matching their expertise to more apt experts, avoiding negative reviews or even malpractice claims. In this paper, we investigate the interplay between reputation and referrals to better understand market outcomes.

To this end, we introduce a dynamic model of strategic referrals among experts. Based on their reputation, experts are approached to solve problems. As problem classification requires costly diagnosis, problems are initially unsorted. Experts treat problems for which their expertise is a good match, and decide whether to treat or refer mismatched problems to more apt experts. An expert solves a problem whenever they treat it successfully, or refer it to another expert who treats it successfully. The expert's reputation then reflects their historic success rate in solving problems.

In the unique steady state of Markov equilibria of our model, competition between experts typically does not result in an efficient matching of problems and experts. This is because the reputation effect, driving future demand, is not powerful enough to perfectly discipline experts' referral behavior. This issue may in fact be exacerbated when experts invest in their abilities. While a social planner wants to increase specialization to grow treatment success and referral rates, experts themselves often prefer to improve their weaker skills and refer less as a result. This is particularly relevant in the light of the recent rise of artificial intelligence in the workplace, allowing professionals to bolster their lesser abilities.

We find that efficient referral outcomes in the market rely on relatively low treatment fees, which potentially discourage specialization in the first place. A more viable way to increase efficiency in the market is for regulation to allow for or even foster referral alliances and partnerships. Even without financial compensation, experts face incentives to enter referral relationships expecting to receive more referrals that match their specialty than outside an alliance. Pervasive small complementary partnerships are incentive-compatible and can restore efficiency in equilibrium. Formally, we introduce a dynamic game with infinitely many agents competing for problems in continuous time. We solve for Markovian equilibria in which each agent conditions their instantaneous rate of referring mismatched problems to other experts on their record, i.e., its thickness (their decaying historic demand) and reputation (their decaying historic success rate). Other experts' records affect an expert's optimal referral rate through aggregate reputation. We establish existence and uniqueness of Markov equilibrium and an associated steady state.

There is a number of possible directions in which our results can be applied or extended. For example, within the realm of referrals, our model could shine light on the growing practice of cross-selling between firms. Alternatively, our framework may lend itself well to analyze scenarios of product quality competition by different types of firms. From a formal perspective, natural next steps are to extend our results to more than two types of experts, or to analyze market outcomes for different concepts of reputation important in alternate settings.

References

- AN, C., A. J. O'MALLEY, AND D. N. ROCKMORE (2018): "Referral paths in the U.S. physician network," *Applied Network Science*, 3(1), 20.
- BAR-ISAAC, H., AND S. TADELIS (2008): "Seller Reputation," *Foundations and Trends in Microeconomics*, 4(4), 273–351.
- BARNETT, M. L., Z. SONG, AND B. E. LANDON (2012): "Trends in Physician Referrals in the United States, 1999-2009," *Archives of Internal Medicine*, 172(2), 163–170.
- BOLTON, P., X. FREIXAS, AND J. SHAPIRO (2007): "Conflicts of interest, information provision, and competition in the financial services industry," *Journal of Financial Economics*, 85(2), 297–330.
- BRYNJOLFSSON, E., D. LI, AND L. R. RAYMOND (2023): "Generative AI at Work," *NBER Working Paper*, p. 31161.
- BUSINESS DEVELOPMENT BANK OF CANADA (2013): "Professional Referral Process," *BDC ViewPoints Study*, https://publications.gc.ca/collections/collection_2017/bdc/Iu133-18-2012-eng.pdf.
- CHANDRA, A., A. FINKELSTEIN, A. SACARNY, AND C. SYVERSEN (2016): "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector," *American Economic Review*, 106(8), 2110–2144.
- CRANDALL, M. J., H. ISHII, AND P. L. LIONS (1992): "User's guide to viscosity solutions of second order partial differential equations," *Bulletin of the American Mathematical Society*, 27(1), 1–67.
- CUI, Z. K., M. DEMIRER, S. JAFFE, L. MUSLOFF, S. PENG, AND T. SALZ (2024): "The Effects of Generative AI on High Skilled Work: Evidence from Three Field Experiments with Software Developers," *SSRN Working Paper*, p. 4945566.
- DAFNY, L. S., K. HO, AND R. S. LEE (2019): "The price effects of cross-market mergers: theory and evidence from the hospital industry," *The RAND Journal of Economics*, 50(2), 286–325.
- DUNN, A., AND A. SHAPIRO (2014): "Do Physicians Possess Market Power," *Journal of Law and Economics*, 57(1), 159–193.
- EFTEKHARI, S., N. YARAGHI, R. D. GOPAL, AND R. RAMESH (2023): "Impact of Health Information Exchange Adoption on Referral Patterns," *Management Science*, 69(3), 1615–1638.
- ELY, J., AND J. VÄLIMÄKI (2003): "Impact of Health Information Exchange Adoption on Referral Patterns," *Management Science*, 69(3), 1615–1638.
- FRANKS, P., G. C. WILLIAMS, J. ZWANZIGER, C. MOONEY, AND M. SORBERO (2000): "Why Do Physicians Vary So Widely in Their Referral Rates?," *Journal of General Internal Medicine*, 15(3), 163–168.

- FRANKS, P., J. ZWANZIGER, C. MOONEY, AND M. SORBERO (1999): "Variations in primary care physician referral rates," *Health Services Research*, 34(1), 323–329.
- FUCHS, W., AND L. GARICANO (2010): "Matching Problems with Expertise in Firms and Markets," *Journal of the European Economic Association*, 8(2/3), 354–364.
- FUCHS, W., L. GARICANO, AND L. RAYO (2015): "Optimal Contracting and the Organization of Knowledge," *The Review of Economic Studies*, 82(2), 632–658.
- FUDENBERG, D., AND D. K. LEVINE (1999): "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica*, 57(4), 759–778.
- GARICANO, L. (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 108(5), 874–904.
- GARICANO, L., AND T. N. HUBBARD (2009): "Specialization, Firms, and Markets: The Division of Labor within and between Law Firms," *Journal of Law, Economics, and Organization*, 25(2), 339–371.
- GARICANO, L., AND T. SANTOS (2004): "Referrals," *American Economic Review*, 94(3), 499–525.
- GODAGER, G., T. IVERSEN, AND A. C.-T. MA (2015): "Do Physicians Possess Market Power," *Journal of Health Economics*, 39, 159–170.
- GONZALEZ, M. L., AND J. A. RIZZO (1991): "Physician Referrals and the Medical Market Place," *Medical Care*, 29(10), 1017–1027.
- GONZALEZ, P. (2010): "Gatekeeping Versus Direct-Access When Patient Information Matters," *Health Economics*, 19(6), 730–754.
- GRASSI, S., AND C.-T. A. MA (2016): "Information acquisition, referral, and organization," *The RAND Journal of Economics*, 47(4), 935–960.
- INDERST, R., AND M. OTTAVIANI (2009): "Misselling through Agents," *American Economic Review*, 99(3), 883–908.
- IVERSEN, T., AND A. C.-T. MA (2011): "Market conditions and general practitioners' referrals," *International Journal of Health Care Finance and Economics*, 11, 245–265.
- LIU, J., AND D. A. HYMAN (2024): "Network analysis of lawyer referral markets: Evidence from Indiana," *Journal of Empirical Legal Studies*, 21(4), 757–785.
- LIU, Q. (2011): "Information Acquisition and Reputation Dynamics," *The Review of Economic Studies*, 78(4), 1400–1425.
- NAM, I., T. S. GRUCA, AND R. TRACY (2010): "The Effects of Competition on Referral Alliances of Professional Service Firms," *Organization Science*, 21(1), 216–231.
- NOY, S., AND W. ZHANG (2023): "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 381(6654), 187–192.

- PARK, I.-U. (2005): “Cheap-Talk referrals of differentiated experts in repeated relationships,” *The RAND Journal of Economics*, 36(2), 391–411.
- RINGBERG, U., N. FLETEN, T. S. DERAAS, T. HASVOLD, AND O. FØRDE (2013): “High referral rates to secondary care by general practitioners in Norway are associated with GPs’ gender and specialist qualifications in family medicine, a study of 4350 consultations,” *BMC Health Services Research*, 13, 147.
- SATTERTHWAITE, M. A. (1979): “Consumer Information, Equilibrium Industry Price, and the Number of Sellers,” *The Bell Journal of Economics*, 10(2), 483–502.
- SHAPIRO, C. (1983): “Premiums for High Quality Products as Returns to Reputations,” *The Quarterly Journal of Economics*, 98(4), 659–680.
- TADELIS, S. (1999): “What’s in a Name? Reputation as a Tradeable Asset,” *American Economic Review*, 88(3), 548–563.
- TIROLE, J. (1996): “A Theory of Collective Reputations (with applications to the persistence of corruption and firm quality),” *The Review of Economic Studies*, 63(1), 1–22.
- VARAS, F., I. MARINOVIC, AND A. SKRZYPACZ (2020): “Random Inspections and Periodic Reviews: Optimal Dynamic Monitoring,” *The Review of Economic Studies*, 87(6), 2893–2937.

A Appendix: Proofs

Proof of Lemma 1.

Define $\underline{m}_j \equiv f(\underline{q}_j, Q_*)$ and $\overline{m}_j \equiv f(\overline{q}_j, Q_*)$. Since $f(\cdot, Q_*)$ is strictly increasing, we know $m^i \in [\underline{m}_j, \overline{m}_j]$ and $q^i \in [\underline{q}_j, \overline{q}_j]$. The weak (viscosity) solution $V^i(\cdot, \cdot)$ uniquely exists on $[\underline{q}_j, \overline{q}_j] \times [\underline{m}_j, \overline{m}_j]$ according to [Crandall, Ishii, and Lions \(1992\)](#). Allowing for differential inclusion, the control r^i is an upper-hemicontinuous correspondence defined on $[\underline{q}_j, \overline{q}_j] \times [\underline{m}_j, \overline{m}_j]$. Since $\frac{dm^i}{dt}$ and $\frac{dq^i}{dt}$ are continuous in r^i , they are both upper-hemicontinuous correspondences. Filippov’s theorem guarantees the existence of solution to the PDE governed by r^i .

For an arbitrary $T > 0$, define a set-valued map Φ_T on the domain $[\underline{q}_j, \overline{q}_j] \times [\underline{m}_j, \overline{m}_j]$:

$$\Phi_T(m^i, q^i) \equiv \left\{ (m_T^i, q_T^i) : \right. \\ \left. (m_t^i, q_t^i) \text{ is a Filippov solution on } [0, T] \text{ given } r_t^i, (m_0^i, q_0^i) = (m^i, q^i) \right\}.$$

$\Phi_T(m^i, q^i)$ is non-empty for all (m^i, q^i) according to Filippov's theorem. $\Phi_T(m^i, q^i)$ is also convex for all (m^i, q^i) because the evolution of (m^i, q^i) is affine in r^i . Finally, Φ_T is upper-hemicontinuous given the continuity of the HJB problem and the optimality condition on r^i . According to Kakutani's fixed point theorem, there exists a fixed point (m_*^i, q_*^i) in the sense that $(m_*^i, q_*^i) \in \Phi_T(m_*^i, q_*^i)$.

Let F_n ($n \geq 0$) be set of fixed points of $\Phi_{T/2^n}$. Then for all n , F_n is non-empty and compact. Moreover, $F_{n+1} \subset F_n$ because the PDE is an autonomous system. Therefore, $F_\infty \equiv \bigcap_{n=0}^\infty F_n$ is non-empty. Note that F_∞ is the set of fixed points of the differential map of the PDE, i.e., (m_*^i, q_*^i) such that $\frac{dm^i}{dt} = 0$ and $\frac{dq^i}{dt} = 0$ for some optimally chosen r_*^i . By definition, (m_*^i, q_*^i, r_*^i) is a fixed point we look for. ■

Proof of Proposition 1.

We use optimal control to show uniqueness of the fixed point. The Hamiltonian reads:

$$e^{-\rho t} f(q_t^i, Q_*) (1 - p_{-j} \pi r_t^i) + \bar{\lambda}_t (1 - r_t^i) + \underline{\lambda}_t r_t^i + \gamma_t^m \phi(f(q_t^i, Q_*) - m_t^i) + \gamma_t^q \phi \left(\frac{f(q_t^i, Q_*)}{m_t^i} (r_t^i \bar{q}_j + (1 - r_t^i) \underline{q}_j - q_t^i) \right),$$

where $\bar{\lambda}_t, \underline{\lambda}_t \geq 0$ are Lagrangian multipliers for the constraint $r_t^i \in [0, 1]$, and γ_t^m, γ_t^q are the costate variables for m^i, q^i . We omit the constant term $\frac{1}{w_j} \int_{\mathcal{I}-j} f(q_*^i, Q_*) p_j \pi r_*^i di'$ from the Hamiltonian because it is outside i 's control.

Optimality requires:

$$f(q_t^i, Q_*) \left(-e^{-\rho t} p_j \pi + \frac{\phi(\bar{q}_j - \underline{q}_j) \gamma_t^q}{m_t^i} \right) = \bar{\lambda}_t - \underline{\lambda}_t, \quad (10)$$

$$-\frac{\phi \gamma_t^q}{(m_t^i)^2} (r_t^i \bar{q}_j + (1 - r_t^i) \underline{q}_j - q_t^i) f(q_t^i, Q_*) - \phi \gamma_t^m = -\frac{d\gamma_t^m}{dt}, \quad (11)$$

$$-\frac{\phi \gamma_t^q}{m_t^i} f(q_t^i, Q_*) + \left(\frac{\phi \gamma_t^q}{m_t^i} (r_t^i \bar{q}_j + (1 - r_t^i) \underline{q}_j - q_t^i) + \phi \gamma_t^m + e^{-\rho t} (1 - p_j \pi r_t^i) \right) f_1(q_t^i, Q_*) = -\frac{d\gamma_t^q}{dt}, \quad (12)$$

with transversality conditions $\lim_{t \rightarrow \infty} \gamma_t^m = \lim_{t \rightarrow \infty} \gamma_t^q = 0$. We look for initial values (m_*^i, q_*^i) such that a constant sequence $(m_t^i, q_t^i) = (m_*^i, q_*^i)$ satisfies the necessary conditions for optimality.

Given $q_*^i \in [\underline{q}_j, \bar{q}_j]$, a steady state requires $m_*^i = f(q_*^i)$ and $r_*^i = \frac{q_*^i - \underline{q}_j}{\bar{q}_j - \underline{q}_j}$. Plugging these into (11)-(12) and applying the transversality conditions, we have $\gamma_t^m = 0$ and $\gamma_t^q = \frac{e^{-\rho t} f_1(q_*^i, Q_*)}{\rho(1+\mu)} (1 - p_{-j} \pi r_*^i)$. Define $h(q, Q) \equiv \frac{f_1(q, Q)}{f(q, Q)}$.

Case 1: $q_*^i = \underline{q}_j$. Then $\bar{\lambda}_t = 0$ and $\underline{\lambda}_t \geq 0$, and (10) requires $(1 - p_{-j} \pi r_*^i) h(q_*^i, Q_*) \frac{\bar{q}_j - \underline{q}_j}{p_{-j}} \frac{\mu}{1+\mu} - \pi \leq 0$ when evaluated at $q_*^i = \underline{q}_j$.

Case 2: $q_*^i = \bar{q}_j$. Then $\bar{\lambda}_t \geq 0$ and $\underline{\lambda}_t = 0$, and (10) requires $(1 - p_{-j} \pi r_*^i) h(q_*^i, Q_*) \frac{\bar{q}_j - \underline{q}_j}{p_{-j}} \frac{\mu}{1+\mu} - \pi \geq 0$ when evaluated at $q_*^i = \bar{q}_j$.

Case 3: $q_*^i \in (\underline{q}_j, \bar{q}_j)$. Then $\bar{\lambda}_t = \underline{\lambda}_t = 0$, and (10) requires $(1 - p_{-j} \pi r_*^i) h(q_*^i, Q_*) \frac{\bar{q}_j - \underline{q}_j}{p_{-j}} \frac{\mu}{1+\mu} - \pi = 0$ when evaluated at q_*^i .

Note that r_*^i decreases in q_*^i and $h(q_*^i, Q_*)$ decreases in q_*^i . Therefore, every parameter combination can be of one and only one case, and the solution $q_*^i \in (\underline{q}_j, \bar{q}_j)$ in Case 3 is unique. Since the existence of a steady state is guaranteed by Lemma 1, this candidate solution must be the true steady state. ■

Proof of Proposition 2.

Define $\underline{Q} \equiv G^{-1}(w_A G(\underline{q}_A) + w_B G(\underline{q}_B)) \geq 0$ and $\bar{Q} \equiv G^{-1}(w_A G(\bar{q}_A) + w_B G(\bar{q}_B)) \leq 1$. Define $h(q, Q) \equiv \frac{f_1(q, Q)}{f(q, Q)}$. For any candidate $Q_* \in [\underline{Q}, \bar{Q}]$, consider $-G(Q_*) + w_A G(q_{A*}(Q_*)) + w_B G(q_{B*}(Q_*))$ as a function of Q_* , where q_{A*} and q_{B*} are implicitly defined in Proposition 1 by (5)-(7). Note:

$$\begin{aligned} -G(Q) + w_A G(q_{A*}(Q)) + w_B G(q_{B*}(Q)) &\geq -G(\underline{Q}) + w_A G(\underline{q}_A) + w_B G(\underline{q}_B) = 0, \\ -G(\bar{Q}) + w_A G(q_{A*}(\bar{Q})) + w_B G(q_{B*}(\bar{Q})) &\leq -G(\bar{Q}) + w_A G(\bar{q}_A) + w_B G(\bar{q}_B) = 0. \end{aligned}$$

By the Intermediate Value Theorem, there exists a Q_* satisfying (8). For uniqueness, take the derivative of the function w.r.t. Q_* :

$$-G'(Q_*) + w_A G'(q_{A*}(Q_*)) q'_{A*}(Q_*) + w_B G'(q_{B*}(Q_*)) q'_{B*}(Q_*), \quad (13)$$

where:

$$q'_{j*}(Q_*) = \begin{cases} \frac{(\bar{q}_j - \underline{q}_j - p_{-j} \pi (q_{j*} - \underline{q}_j)) h_2(q_{j*}, Q_*)}{p_{-j} \pi h(q_{j*}, Q_*) - (\bar{q}_j - \underline{q}_j - p_{-j} \pi (q_{j*} - \underline{q}_j)) h_1(q_{j*}, Q_*)} & \text{if } q_{j*} \in (\underline{q}_j, \bar{q}_j), \\ 0 & \text{if } q_{j*} = \underline{q}_j \text{ or } q_{j*} = \bar{q}_j, \end{cases}$$

when q'_{j*} exists. Note that $q'_{j*}(Q_*) \geq 0$ by [Assumption 1](#). Also, since $h_2(q_{j*}, Q_*)$ is small, so is $q'_{j*}(Q_*)$. Since $G'(Q_*)$ is strongly positive on $[Q, \bar{Q}]$, we know (13) is negative. Therefore, there exists a unique Q_* in the general equilibrium. ■

Proof of [Proposition 3](#).

Wlog, consider a marginal increase of η_A^A . Define $h(q, Q) \equiv \frac{f_1(q, Q)}{f(q, Q)}$.

If $(1 - p_B \pi r_{A*})h(q_{A*}, Q_*) \frac{\bar{q}_A - q_A}{p_B} \frac{\mu}{1+\mu} - \pi \neq 0$, then $\frac{dr_{A*}}{d\eta_A^A} = 0$, and $q_{A*} = r_{A*}\bar{q}_A + (1 - r_{A*})\underline{q}_A$ strictly increases in η_A^A . If $(1 - p_B \pi r_{A*})h(q_{A*}, Q_*) \frac{\bar{q}_A - q_A}{p_B} \frac{\mu}{1+\mu} - \pi = 0$, then the Implicit Function Theorem yields:

$$\frac{dq_{A*}}{d\eta_A^A} = \frac{p_A p_B \pi r_{A*} h(q_{A*}, Q_*) + (q_{A*} - \underline{q}_A)(1 - p_B \pi r_{A*}) \frac{dq_{B*}}{d\eta_A^A} \frac{\partial Q_*}{\partial q_{B*}} h_2(q_{A*}, Q_*)}{p_B \pi r_{A*} h(q_{A*}, Q_*) - (q_{A*} - \underline{q}_A)(1 - p_B \pi r_{A*}) \left(\frac{\partial Q_*}{\partial q_{A*}} h_2(q_{A*}, Q_*) + h_1(q_{A*}, Q_*) \right)}.$$

With [Assumption 1\(ii\)](#), $h_2(q_{A*}, Q_*)$ is negligible relative to other terms, and therefore $\frac{dq_{A*}}{d\eta_A^A} > 0$ whenever q_{A*} is differentiable.

If $(1 - p_A \pi r_{B*})h(q_{B*}, Q_*) \frac{\bar{q}_B - q_B}{p_A} \frac{\mu}{1+\mu} - \pi \neq 0$, then $\frac{dr_{B*}}{d\eta_A^A} = 0$, and $q_{B*} = r_{B*}\bar{q}_B + (1 - r_{B*})\underline{q}_B$ weakly increases in η_A^A . If $(1 - p_A \pi r_{B*})h(q_{B*}, Q_*) \frac{\bar{q}_B - q_B}{p_A} \frac{\mu}{1+\mu} - \pi = 0$, then the Implicit Function Theorem yields:

$$\frac{dq_{B*}}{d\eta_A^A} = \frac{p_A r_{B*} h(q_{B*}, Q_*) + (q_{B*} - \underline{q}_B)(1 - p_A \pi r_{B*}) \frac{dq_{A*}}{d\eta_A^A} \frac{\partial Q_*}{\partial q_{A*}} h_2(q_{B*}, Q_*)}{p_A \pi r_{B*} h(q_{B*}, Q_*) - (q_{B*} - \underline{q}_B)(1 - p_A \pi r_{B*}) \left(\frac{\partial Q_*}{\partial q_{B*}} h_2(q_{B*}, Q_*) + h_1(q_{B*}, Q_*) \right)}.$$

With [Assumption 1\(ii\)](#), $h_2(q_{B*}, Q_*)$ is negligible relative to other terms, and therefore $\frac{dq_{B*}}{d\eta_A^A} > 0$ whenever q_{B*} is differentiable. ■

Proof of [Proposition 4](#).

Wlog, consider a marginal increase of η_A^B . Define $h(q, Q) \equiv \frac{f_1(q, Q)}{f(q, Q)}$.

(i) If $(1 - p_B \pi r_{A*})h(q_{A*}, Q_*) \frac{\bar{q}_A - q_A}{p_B} \frac{\mu}{1+\mu} - \pi \neq 0$, then $r_{A*} = 0$ or $r_{A*} = 1$, $\frac{dr_{A*}}{d\eta_A^B} = 0$, and $q_{A*} = r_{A*}\bar{q}_A + (1 - r_{A*})\underline{q}_A$ stays constant in η_A^B if $r_{A*} = 1$ and strictly increases in η_A^B if $r_{A*} = 0$. If $(1 - p_B \pi r_{A*})h(q_{A*}, Q_*) \frac{\bar{q}_A - q_A}{p_B} \frac{\mu}{1+\mu} - \pi = 0$, then the Implicit Function Theorem yields:

$$\frac{dq_{A*}}{d\eta_A^B} = - \frac{p_B(1 - p_B \pi) r_{A*} h(q_{A*}, Q_*) - (q_{A*} - \underline{q}_A)(1 - p_B \pi r_{A*}) \frac{dq_{B*}}{d\eta_A^B} \frac{\partial Q_*}{\partial q_{B*}} h_2(q_{A*}, Q_*)}{p_B \pi r_{A*} h(q_{A*}, Q_*) - (q_{A*} - \underline{q}_A)(1 - p_B \pi r_{A*}) (h_1(q_{A*}, Q_*) + \frac{\partial Q_*}{\partial q_{A*}} h_2(q_{A*}, Q_*))}.$$

With [Assumption 1\(ii\)](#), $h_2(q_{A^*}, Q_*)$ is negligible relative to other terms, and therefore $\frac{dq_{A^*}}{d\eta_A^B} < 0$ whenever q_{A^*} is differentiable. Since $r_{A^*} = \frac{q_{A^*} - \underline{q}_A}{\bar{q}_A - \underline{q}_A}$ where \underline{q}_A increases in η_A^B , we know r_{A^*} decreases in η_A^B .

(ii) If $(1 - p_A \pi r_{B^*})h(q_{B^*}, Q_*) \frac{\bar{q}_B - \underline{q}_B}{p_A} \frac{\mu}{1+\mu} - \pi \neq 0$, then $r_{B^*} = 0$ or $r_{B^*} = 1$, $\frac{dr_{B^*}}{d\eta_A^B} = 0$, and $\frac{dq_{B^*}}{d\eta_A^B} = 0$. If $(1 - p_B \pi r_{A^*})h(q_{A^*}, Q_*) \frac{\bar{q}_A - \underline{q}_A}{p_B} \frac{\mu}{1+\mu} - \pi = 0$, then the Implicit Function Theorem yields:

$$\frac{dq_{B^*}}{d\eta_A^B} = \frac{(q_{B^*} - \underline{q}_B)(1 - p_A \pi r_{B^*}) \frac{dq_{A^*}}{d\eta_A^B} \frac{\partial Q_*}{\partial q_{A^*}} h_2(q_{B^*}, Q_*)}{p_A \pi r_{B^*} h(q_{B^*}, Q_*) - (q_{B^*} - \underline{q}_B)(1 - p_A \pi r_{B^*})(h_1(q_{B^*}, Q_*) + \frac{\partial Q_*}{\partial q_{B^*}} h_2(q_{B^*}, Q_*))}.$$

With [Assumption 1\(ii\)](#), $h_2(q_{B^*}, Q_*)$ in the denominator is negligible relative to other terms, but $h_2(q_{B^*}, Q_*) \geq 0$ in the numerator determines the sign of $\frac{dq_{B^*}}{d\eta_A^B}$, which is the same as $\frac{dq_{A^*}}{d\eta_A^B}$. Since $r_{B^*} = \frac{q_{B^*} - \underline{q}_B}{\bar{q}_B - \underline{q}_B}$, we know r_{B^*} moves in the same direction as q_{B^*} . ■

Proof of [Proposition 5](#).

Take the limit as $w_A = 1 - w_B \rightarrow 1$. By continuity of G , in the limit we have $\frac{\partial Q_*}{\partial q_{A^*}} = 1$ and $\frac{\partial Q_*}{\partial q_{B^*}} = 0$. Also, from the proof of [Proposition 3](#) and [Proposition 4](#), we know:

$$\begin{aligned} \frac{\partial q_{A^*}}{\partial \eta_A^A} &= \frac{p_A p_B \pi r_{A^*} h(q_{A^*}, Q_*)}{p_B \pi r_{A^*} h(q_{A^*}, Q_*) - (q_{A^*} - \underline{q}_A)(1 - p_B \pi r_{A^*})(h_1(q_{A^*}, Q_*) + h_2(q_{A^*}, Q_*))}, \\ \frac{\partial q_{A^*}}{\partial \eta_A^B} &= \frac{p_B(1 - p_B \pi) r_{A^*} h(q_{A^*}, Q_*)}{p_B \pi r_{A^*} h(q_{A^*}, Q_*) - (q_{A^*} - \underline{q}_A)(1 - p_B \pi r_{A^*})(h_1(q_{A^*}, Q_*) + h_2(q_{A^*}, Q_*))}. \end{aligned}$$

With these, we can take the derivative of expert type A 's equilibrium payoff $f(q_{A^*}, Q_*)(1 - p_B \pi r_{A^*}) + \frac{w_B}{w_A} f(q_{B^*}, Q_*) p_A \pi r_{B^*}$ w.r.t. η_A^B and η_A^A respectively, and then take the difference between them. In the limit as $w_A = 1 - w_B \rightarrow 1$, $f_1(q_{A^*}, q_{A^*}) + f_2(q_{A^*}, q_{A^*}) \rightarrow 0$ and $h_1(q_{A^*}, q_{A^*}) + h_2(q_{A^*}, q_{A^*}) \rightarrow 0$, we find the difference to be:

$$\frac{(1 - p_B \pi r_{A^*}) f(q_{A^*}, q_{A^*})}{\eta_B^B - \eta_A^B} > 0.$$

By continuity of the system, this difference remains positive when w_A is sufficiently close to 1, and $|f_1(q_{A^*}, q_{A^*}) + f_2(q_{A^*}, q_{A^*})|$ and $|h_1(q_{A^*}, q_{A^*}) + h_2(q_{A^*}, q_{A^*})|$ are sufficiently small. ■

Proof of [Proposition 6](#).

Consider $\left(1 - p_{-j}\pi \frac{q_{j^*} - \underline{q}_j}{\bar{q}_j - \underline{q}_j}\right) h(q_{j^*}, Q_*) \frac{\bar{q}_j - \underline{q}_j}{p_{-j}} \frac{\mu}{1+\mu} - \pi$, the left-hand side of (10). Noting Q_* as a function of q_{A^*} and q_{B^*} , $\left(1 - p_{-j}\pi \frac{q_{j^*} - \underline{q}_j}{\bar{q}_j - \underline{q}_j}\right) h(q_{j^*}, Q_*)$ decreases in q_{j^*} because of [Assumption 1](#).

When μ increases or when $h(\cdot, Q_*)$ increases pointwise as a function, the left-hand side of (10) increases as well. Either q_{j^*} remains at \bar{q}_j or \underline{q}_j as a corner solution, or q_{j^*} has to increase to keep (10) satisfied. $r_{j^*} = \frac{q_{j^*} - \underline{q}_j}{\bar{q}_j - \underline{q}_j}$ moves in the same direction as q_{j^*} . ■

Proof of [Proposition 7](#).

Consider $\left(1 - p_{-j}\pi \frac{q_{j^*} - \underline{q}_j}{\bar{q}_j - \underline{q}_j}\right) h(q_{j^*}, Q_*) \frac{\bar{q}_j - \underline{q}_j}{p_{-j}} \frac{\mu}{1+\mu} - \pi$, the left-hand side of (10). Noting Q_* as a function of q_{A^*} and q_{B^*} , $\left(1 - p_{-j}\pi \frac{q_{j^*} - \underline{q}_j}{\bar{q}_j - \underline{q}_j}\right) h(q_{j^*}, Q_*)$ decreases in q_{j^*} because of [Assumption 1](#).

When π increases, the left-hand side of (10) decreases. Either q_{j^*} remains at \bar{q}_j or \underline{q}_j as a corner solution, or q_{j^*} has to decrease to keep (10) satisfied. $r_{j^*} = \frac{q_{j^*} - \underline{q}_j}{\bar{q}_j - \underline{q}_j}$ moves in the same direction as q_{j^*} . ■

Proof of [Lemma 2](#).

In the proof of [Proposition 1](#), revise the first term of the Hamiltonian to $e^{-\rho t} f(q_t^i, Q_*) (1 - p_{-j}\pi(1 - y_{-j})r_t^i)$. The problem is the same as in the main model except π is replaced by $\pi(1 - y_{-j})$. According to (10), $r_{A^*} = r_{B^*} = 1$ if and only if $(1 - p_{-j}\pi(1 - y_{-j})r_{j^*})h(q_{j^*}, Q_*) \frac{\bar{q}_j - \underline{q}_j}{p_{-j}} \frac{\mu}{1+\mu} - \pi(1 - y_{-j}) \geq 0$ evaluated at $r_* = 1$, $q_{j^*} = \bar{q}_j$ and $Q_* = \bar{Q}$ for $j = A, B$, or equivalently, (9) holds.

Noting Q_* as a function of q_{A^*} and q_{B^*} , $\left(\frac{1}{1-y_{-j}} - p_{-j}\pi \frac{q_{j^*} - \underline{q}_j}{\bar{q}_j - \underline{q}_j}\right) h(q_{j^*}, Q_*)$ decreases in q_{j^*} because of [Assumption 1](#). It also increases in y_{-j} . With a higher y_{-j} , either q_{j^*} remains at \bar{q}_j or \underline{q}_j as a corner solution, or q_{j^*} has to increase to keep (10) satisfied. $r_{j^*} = \frac{q_{j^*} - \underline{q}_j}{\bar{q}_j - \underline{q}_j}$ moves in the same direction as q_{j^*} . ■

Proof of [Proposition 8](#).

Before forming an alliance, the steady state payoff of type j is $f(q_{j^*}, Q_*) (1 - p_{-j}\pi r_{j^*}) + W$, where W is the payoff from the referral from other experts, depending on the current structure of alliances in the economy. After forming the referral alliance of n_j and n_{-j} experts, the aggregate reputation Q_* and the behavior of the experts

outside do not change. The payoff of type j becomes

$$f(\tilde{q}_{j*}, Q_*)(1 - p_{-j}\pi\tilde{r}_{j*}(1 - y_{-j})) + \frac{n_{-j}}{n_j}f(\tilde{q}_{-j*}, Q_*)p_j\pi\tilde{r}_{-j*}(1 - y_j) + W,$$

where \tilde{q}_{j*} and \tilde{r}_{j*} indicate the new reputation and referral rate of type j in the alliance.

Note that the last term, W , remains the same as before entering the alliance. The second term, $\frac{n_{-j}}{n_j}f(\tilde{q}_{-j*}, Q_*)p_j\pi\tilde{r}_{-j*}(1 - y_j)$, is a result of the contract and experts strictly prefer referring within the alliance. This term is non-negative. The first term is a continuous function of y_{-j} . When $y_{-j} = 0$, we have $\tilde{q}_{j*} = q_{j*}$ and $\tilde{r}_{j*} = r_{j*}$, and the payoff is higher than before entering the alliance due to the second term. When $y_{-j} > 0$, $\frac{d}{dy_{-j}}f(\tilde{q}_{j*}, Q_*)(1 - p_{-j}\pi\tilde{r}_{j*}(1 - y_{-j})) = p_{-j}\pi\tilde{r}_{j*} > 0$ if \tilde{q}_{j*} is a corner solution, and

$$\begin{aligned} \frac{d}{dy_{-j}}f(\tilde{q}_{j*}, Q_*)(1 - p_{-j}\pi\tilde{r}_{j*}(1 - y_{-j})) &= p_{-j}\pi\tilde{r}_{j*} \\ &+ \frac{h(\tilde{q}_{j*}, Q_*)p_{-j}\pi/\mu}{p_{-j}\pi(1 - y_{-j})h(\tilde{q}_{j*}, Q_*) - (\bar{q}_j - \underline{q}_j)(1 - p_{-j}\pi(1 - y_{-j})\tilde{r}_{j*})h_1(\tilde{q}_{j*}, Q_*)} > 0 \end{aligned}$$

if \tilde{q}_{j*} is interior. ■

Proof of Proposition 9.

Given any aggregate reputation Q , note that the total payoff within a partnership is simply $n_A f(q_{A*}, Q) + n_B f(q_{B*}, Q)$, because the total fee of 1 generated by a problem goes to the partnership, and there is no referral from outside the partnership. Since Q is treated as fixed, the partnership's total payoff is maximized when $q_{A*} = \bar{q}_A$ and $q_{B*} = \bar{q}_B$, i.e., full referral. Because utility is transferable within the partnership, every Pareto-efficient contract must maximize total payoff, and hence guarantees full referral. Moreover, it is easy to see that such a Pareto-efficient contract exists.

If any finite number of experts jointly deviate from the pervasive partnership arrangement, they can try to quit their respective partnerships, form a new one, join in existing ones, or simply change the terms of contract. These deviations involve only a finite number of partnerships, imposing no effect on the aggregate reputation Q . Consider the total payoff of all experts in those affected partnerships. Since in each partnership the contract is already Pareto-efficient, it remains Pareto-efficient

for all these partnerships jointly. Therefore, there exists no alternative partnership formation with new contracts that constitutes a Pareto improvement. The pervasive partnership structure is thus stable against joint deviations y finitely many experts.

■