DEPARTMENT OF ECONOMICS
JOHANNES KEPLER UNIVERSITY OF
LINZ

Experts vs. Discounters:
Consumer Free Riding and Experts Withholding Advice in Markets for Credence Goods

by

Uwe Dulleck*)
Rudolf Kerschbamer

Working Paper No. 0509
December 2005

Johannes Kepler University of Linz
Department of Economics
Altenberger Strasse 69
A-4040 Linz - Auhof, Austria
www.econ.jku.at

*)uwe.dulleck@jku.at
phone +43 (0)70 2468 -8239, -8238 (fax)

Experts vs. Discounters:

Consumer Free Riding and Experts Withholding Advice in Markets for Credence Goods

Uwe Dulleck* and Rudolf Kerschbamer†

September 2005

Abstract

This paper studies price competition between experts and discounters in a market for credence goods. While experts can identify a consumer's problem by exerting costly but unobservable diagnosis effort, discounters just sell treatments without giving any advice. The unobservability of diagnosis effort induces experts to use their tariffs as signaling devices. This makes them vulnerable to competition by discounters. We explore the conditions under which experts survive competition by discounters and find that there exist situations in which adding a single customer to a large population of existing consumers leads to a switch from an experts only to a discounters only market. We also discuss whether vertical restraints can alleviate these inefficiencies.

JEL Classifications L15, D82, D40

Keywords Experts, Discounters, Credence Goods, Vertical Restraints

*Department of Economics, University of Vienna, Hohenstaufengasse 9, A - 1010 Vienna, E-Mail uwe.dulleck@univie.ac.at, Tel. +43-1-4277 374 27, Fax +43-1-4277 9 374

†Department of Economics, University of Innsbruck and CEPR, E-Mail rudolf.kerschbamer@uibk.ac.at, Tel. +43-512-507 7400, Fax +43-512-507 2980

1 Introduction

Suppose your washing machine leaks water. The cause of the loss of water could either be a minor problem (a seal is porous and needs to be replaced) or a major problem (the drum is rust-eaten; the best advice is to buy a new machine). You only know that the machine leaks, and even if you are able to replace a porous seal, it is hard for you to determine what seal to replace and whether replacing a seal is sufficient to fix the problem or not. What are your options? You could go to a hardware store and buy either a seal or a new machine. Alternatively, you could call a repair service to have a look at the machine. If you contact the firm, it may either send you an experienced technician (at high cost), or a bungler (at low cost). You are not able to distinguish a real expert from a bungler. The former is able to determine whether replacing a seal will do, the bungler doesn't know much more than you and will make a blind diagnosis. After hearing the advice and the cost of the suggested treatment, you can issue the repair or, with some excuse, turn down the offer. For instance, if the repair person recommends to buy a new machine, you could ask for some time to think about that, and then go to the next warehouse to buy a new machine at a lower price. Or, if he recommends some minor repair which you deem too expensive, you could argue that you are still thinking of buying a new machine and then go to the hardware store to buy a new seal and conduct the recommended treatment yourself.

Similar situations are ubiquitous. For instance, when your car's ignition doesn't work you can go to a backyard garage and ask to replace the battery or the generator; alternatively, you can visit a mechanic who is able to identify which maintenance needs to be done by exerting costly (but unobservable) effort. Once the diagnosis has been made, you can again either issue the repair or, with some excuse, turn down the offer. Similarly, one may buy a given PC either in a supermarket or from some expert seller, who can (at some cost) identify ones exact needs. All these examples have in common that the consumer feels a need but cannot tell which type of good or service meets his need best. He can blindly buy some treatment from a discounter, or, he can visit an expert. Experts are able to identify the treatment that fits a consumer's need best by incurring a diagnosis cost, but, they face the risk that once the advice is provided, the customer turns away and buys what he needs at a cheap place around the corner.

Goods and services where an expert knows more about the quality a consumer needs than the consumer himself are called credence goods. In the literature on credence goods (see Dulleck and Kerschbamer 2005a for a survey) most contributions ignore consumers' option to free ride on a given advice. This is done by either assuming that diagnosis needs no special effort (cf., e.g. Pitchik and Schotter 1987, and Sülzle and Wambach 2005) or that diagnosis effort is observable and verifiable so that a (fair) diagnosis fee can be imposed on the consumer (see, for instance, Wolinsky 1993 and 1995, or Emons 1997 and 2001). In this article, we study the incentives for experts to invest effort in diagnosis if diagnosis effort is both costly and unobservable, and if they face competition by discounters who are not able to perform a diagnosis. We show that the existence of discounters can unravel a market that would otherwise be efficiently served by experts who invest in diagnosis. In case of unraveling experts stop investing diagnosis effort because customers would free-ride on their advice and then buy what is necessary from the next discount outlet.

The basic features of our model are as follows On the demand side, there are many consumers in the market. Each consumer has either a (minor) problem requiring a cheap treatment $\underline{c}$, or a (major) problem requiring an expensive treatment $\overline{c}$. The customer knows that he has a problem, but does not know which one.

On the supply side of the market there are two types of treatment providers experts and discounters. The distinction between the two types is not in the range of treatments they provide, but only in their ability to determine a consumer's need. A discounter is unable to determine a consumer's need. She just offers a menu of treatments from which consumers have to choose.

Experts, on the other hand, are able to identify the quality that fits a consumer's need best by incurring a diagnosis cost. The consumer does not observe whether or not the expert incurred the cost. He only hears the expert's recommendation. After learning the recommendation, the consumer either buys the recommended treatment at the price the expert asks for it, or he visits another treatment provider. Second visits are costly (and inefficient), however, because consumers incur a search (or switching) cost for each provider they visit.

An important question with unobservable diagnosis effort is whether experts can signal their diagnosis effort through their choice of diagnosis and treatment prices as well as through their choice of warranty payments (for the case of treatment failure). The answer turns out to be yes, but at the cost of being vulnerable to competition by discounters. The reason is as follows With diagnosis effort unobservable, experts must be prevented from choosing one of the following two cheating strategies abstaining from diagnosis and potentially undertreating the consumer (that is, blindly recommending a cheap, low quality treatment), and abstaining from diagnosis and potentially overtreating the consumer (that is, blindly recommending an expensive, high quality treatment). Since the final success of service is observable and verifiable in our model, the undertreating incentive is easily removed by experts offering a warranty for the case of treatment failure. To remove the overtreating incentive, charging the diagnosis fee without providing one and selling the expensive, high quality treatment for sure must be unprofitable. This is only possible if the mark-up on high quality treatments is set to zero and if the diagnosis is given for free.

The necessity to sell high quality without a mark-up and provide diagnosis free of charge implies that diagnosis costs must be earned only through the mark-up on minor treatments. This leads to two different kinds of problem.

First, to a free rider problem If the cost of visiting a second provider is low then discounters are able to attract consumers who have learned from an expert that they need a low quality treatment. If all consumers have low search costs then experts are unable to survive as full service providers on the market. But even if almost all consumers have high search cost the experts' market might be cannibalized by discounters. To see the problem, consider a market in which all consumers should efficiently visit an expert and in which only one consumer has low, the rest high search cost. Then the low search cost consumer has an incentive to consult an expert to get a free diagnosis and to switch to a discounter if the expert recommends the minor treatment. But if this customer switches after receiving the free diagnosis then experts must increase the mark-up on the minor intervention to finance the free-riding customer's diagnosis effort. This price-increase might lead even more customers to free ride on the diagnosis effort. As our analysis reveals the resulting domino effect might lead to a complete unraveling of the experts' market.

Second, if consumers differ in their expected cost of efficient treatment then an adverse selection problem arises. The reason is, that the price structure chosen for signaling reasons implies a cross-subsidization of consumers who are likely to need a major intervention (high cost consumers) by consumers who are likely to need a minor intervention (low cost consumers). This cross-subsidization invites (all) high cost consumers who efficiently should buy an expensive treatment from a discounter to consult an expert for diagnosis. At the same time (some) low cost consumers who efficiently should visit an expert for diagnosis buy blindly the cheap treatment from a discounter. Anticipating this adverse selection problem, experts increase the mark-up on the minor intervention to avoid losses. This price-increase might again lead to a chain-reaction resulting in a complete unraveling of the experts' market.

Our analysis is related to several strands of previous literature. First, to the literature on credence goods. The credence goods paper closest to ours is Pesendorfer and Wolinsky (2003). As in the present paper they consider a market in which an expert must exert costly but unobservable effort to identify the service that meets a consumer's needs best. Their main focus is on the role of a specific mechanism – the gathering of multiple opinions

– in disciplining experts' behavior. A crucial assumption in the Pesendorfer/Wolinsky analysis is that the final success of service is not contractible. Otherwise, the incentive problem stemming from the unobservability of diagnosis effort could easily be solved by an appropriate choice of diagnosis and treatment prices as well as of warranty payments for the case of treatment failure. In contrast, in our model the success of treatment is observable and verifiable and the problem analyzed here stems from the existence of discounters who cannibalize the experts' market.

Our analysis is also related to the papers by Bouckaert and Degryse (2000) and Emons (2000) on competition between safe and risky experts. In these articles consumers face the choice between visiting an expensive expert directly and first trying to solve the problem using a cheap expert. While the expensive expert can solve the problems of all consumers, the cheap expert's repair technology is not always successful. If the cheap risky expert fails, a consumer ends up with the expensive safe expert paying for the service twice. There are several distinctions between these two papers and the setting considered here. First of all, these papers abstract completely from both, experts' incentive to provide a serious diagnosis and their incentive to provide the appropriate treatment. Also, there is no other asymmetric information involved in the models; that is, consumers and producers have exactly the same information about the magnitude of consumers' search cost and their probability of success at the two stores. Finally, in contrast to the setting considered here, this literature also abstracts from the possibility of warranties for cheap sellers. Thus, when translated to the language of the present paper, this literature studies price competition between two discounters, one selling only the cheap treatment $\underline{c}$, the other selling only the expensive treatment $\overline{c}$.

Another related paper is Glazer and McGuire (1996). The basic setup is similar to the one studied in the literature on competition between safe and risky experts, the main difference being that in the Glazer and McGuire paper consumers do not know their success probability with the risky seller.

The risky seller learns this probability by diagnosing the consumer. He then decides whether to refer the consumer to the safe seller. The focus of this paper is on the question of whether in equilibrium there is socially optimal referral from the risky to the safe seller. As in a large part of the credence goods literature, diagnosing a consumer is assumed to be costless in Glazer and McGuire (1996), so the issues studied in the present paper don't arise.

One of the problems analyzed in the present paper (free-riding on experts' advice) has close parallels in the literature on vertical restraints and retail price maintenance (RPM). The classical RPM literature (the seminal paper is Telser 1960; other entries include Marvel and McCafferty 1984, Klein and Murphy 1988, and Shaw 1994) studies situations in which sales at the retail level depend both on retail prices and on the amount of "special services" the retailers provide jointly with the product. Examples for such services are test drives, pre-sale demonstrations of the product, or, in the case of Marvel and McCafferty (1984), a certification of the quality or stylishness of the product. Since these services have a public good characteristic in that the service provided by one retailer also benefits consumers who purchase from other sellers, retailers who do not provide the special service can get a free ride at the expense of those who have convinced consumers to buy the product. As a consequence, none of the retailers has an incentive to offer the special service. In this situation RPM, used as a price floor, can alleviate the problem because it prevents price competition and channels competition into non-price dimensions such as service. The present paper can be seen as complementary to the existing RPM literature in that it provides (i) a new motivation for the use of RPM (in the traditional RPM literature, the special service consists of demonstration or certification activities for a homogeneous product; by contrast, in the present context there are different types or qualities of a good or service and the special service consist in helping the consumer to identify the quality that fits his needs best), and (ii) a new formalization of the special-service free-rider story which is more in line with the original Telser argument envisioning competition between retailers providing special services and charging high prices and retailers providing no service and charging low prices (in the existing formal literature on RPM there is only one type of retailer and the problem is to induce this type of retailer to provide the desired service[1]). Although the driving forces are similar, the implications of our analysis differ in important aspects from those in the traditional RPM literature. For instance, both Telser's special services theory and Marvel and McCafferty's quality certification theory would predict that RPM is used for products which are unfamiliar to the mass of consumers and that RPM usage declines as the good or item becomes better known. In contrast, in the present context it is not the product that is unknown to consumers, it is their own condition; so, if RPM is helpful in the present context, then it is helpful irrespective of whether the products are familiar to the consumers or not. Also, in the traditional RPM literature, high quality products are typical candidates for RPM (Telser 1960, p.95; Marvel and McCafferty 1984, p. 347f). By contrast, in the present setting RPM is potentially helpful for low quality treatments while it is definitively harmful when imposed on high quality treatments. We will discuss at the end of our analysis whether, and if yes, how RPM and other instruments of vertical restraints can help to alleviate the inefficiencies in the present credence goods setting (see Section 6).

The rest of the paper is organized as follows. Section 2 presents our basic model of competition between experts and non-experts. In Section 3 we characterize the efficient diagnosis and treatment policy and then show that the efficient solution could be sustained in equilibrium if experts' diagnosis effort was observable and verifiable. Then we turn to our model with unobservable diagnosis effort. Section 4 characterizes the inefficiencies in the homogeneous customers case. In Section 5 we show that the inefficiencies of the homogeneous consumers case amplify if consumers are heterogeneous.

[1]An exception is Bolton and Bonano (1988). The situation studied there is quite different, however, since consumers are assumed to be able to benefit from a given retailer's services only if they purchase the good from him. Thus, free-riding in the provision of costly services is not an issue there.

Section 6 discusses instruments of vertical restraints to solve the free-rider- and adverse-selection-problems in our framework. Section 7 concludes.

2 The Basic Model

With credence goods, consumers are never sure which quality of the good or service they actually need. To model this situation we assume that each consumer (he) has either a major or a minor problem. The customer knows that he has a problem, but does not know how severe it is. He only knows that he has an *ex ante* probability of h that he has the major problem and a probability of $(1-h)$ that he has the minor one. The major problem requires an expensive treatment $\overline{c}$, the minor problem requires a cheap treatment $\underline{c}$. The cost of the expensive treatment is $\overline{c}$ and the cost of the cheap treatment is $\underline{c}$, with $\overline{c} > \underline{c}$.[2] The expensive treatment fixes either problem while the cheap one is only good for the minor problem.

Table 1 represents the per period utility of a consumer given the type of treatment he needs and the type he gets. If the type of treatment is sufficient, a consumer gets utility v. Otherwise he gets 0. To motivate this payoff structure consider the washing machine example introduced earlier. The machine may have either a minor problem (a seal is porous and needs to be replaced) or a major problem (the drum is rust-eaten; the best advice is to buy a new machine), with the outcomes being 'washing machine works correctly' (if appropriately treated or overtreated) and 'machine is still leaking' (if undertreated). The case of undertreatment is the upper right cell of the table, the case of overtreatment is the lower left cell. Note that overtreatment is not detected by the customer ($v = v$) and hence cannot be ruled out by institutional arrangements. This is not the case with undertreatment; it is detected by the customer ($0 < v$) and might even be verifiable. In the present paper we assume that this is the case. This means that payments can be conditioned on the resolution of the problem. We also assume that

[2]For convenience, both the type of treatment and the associated cost is denoted by c.

Customer's utility		Customer needs	
		$\underline{c}$	$\overline{c}$
Customer	$\underline{c}$	v	0
gets	$\overline{c}$	v	v

Table 1: Utility from a Credence Good

the type of treatment is observable and verifiable so that payments can also be conditioned on the type of treatment.

Let us now describe the market environment. On the demand side there is a continuum of mass one of consumers with the above characteristics. Each consumer can visit one ore more treatment providers. The consumer incurs a search cost s per provider he samples, independently of whether or not he chooses to be treated by this provider. This cost represents the time and effort incurred in searching for a provider. As will become clear below, the variable s can also be interpreted as the remorse felt by a consumer if he decides to visit a second seller after having got an advice from the first one.

On the supply side there are two types of treatment providers, experts and discounters. In both sub-markets, the one for experts and the one for discounters, there are at least two sellers. Each seller (she) can serve arbitrarily many consumers. The distinction between the two types of sellers is not in the spectrum of treatments they provide, but only in the ability to determine a consumer's need. A discounter is unable to determine a consumer's need. She just offers a menu of treatments and consumers have to choose themselves. An expert, on the other hand, can identify the type of treatment the consumer needs by incurring a diagnosis cost c. The consumer does not observe whether or not the expert incurred the cost. He only learns the expert's recommendation.

The interaction between consumers and treatment providers is modelled as follows. Time is divided into two periods. Before the first period begins, experts and discounters simultaneously announce their tariffs. A tariff by a discounter specifies a price $\underline{q}$ for the minor treatment and a price $\overline{q}$ for the major one. A tariff by an expert specifies a diagnosis fee p for the recommendation, a price $\underline{p}$ for the minor treatment and a price $\overline{p}$ for the major one. An expert's tariff might also specify a transfer payment t for the case of treatment failure. At the beginning of period 1 consumers enter the market and – upon observing the tariffs available in the market – each consumer decides which provider (if any) he visits. When a consumer visits a discounter, he specifies which kind of treatment he wants. The discounter then provides the treatment and charges the price posted for it. When a consumer consults an expert, he has to pay the diagnosis fee p in advance. In exchange, the expert makes a recommendation. The consumer doesn't observe whether the expert's recommendation is based on a serious diagnosis at cost c or not. After learning the recommended treatment, the consumer decides whether to receive it. If he refuses the treatment, he either leaves the market or continues to search for another service provider by spending another search cost s. If the consumer accepts, the expert provides the recommended treatment at the price specified for this service. The first period ends with each consumer having either left the market or bought a treatment. If the treatment a consumer got is sufficient to solve his problem he leaves the market. Otherwise he loses v in this period and either buys $\overline{c}$ from the same provider or continues search in the second period. If a consumer's problem is left untreated for two periods, it becomes irreparable and the consumer leaves the market. There is no discounting.

Consumers are minimizers of expected cost. The total cost to a consumer who visited n $(= 1, 2, 3, ..)$ different providers and got a sufficient treatment in period r $(= 1, 2, 3;$ period 3 here stands for the case where the consumer's problem is left untreated for two periods[3]) is $ns + (r - 1)v$ plus the sum of diagnosis and treatment prices paid in the course of his search, minus

[3] As is easily verified, our analysis and results would remain unaffected if we assumed instead that $r \in \{1, 2, x\}$, where period $x \geq 3$ stands for the case where the consumer's problem is left untreated for two periods.

possible transfers for insufficient treatments. By assumption, if a consumer is indifferent between visiting a service provider and not visiting a service provider, he decides for a visit. Also, if a customer who decides for a visit is indifferent between visiting an expert and visiting a discounter, he decides for the expert and if he is indifferent between two or more experts (or two or more discounters), he randomizes (with equal probability) among them.

Treatment providers maximize expected profit. The profit a discounter derives from a customer who visited her is simply the price of the treatment sold minus treatment cost. The profit an expert derives from a customer depends on whether she incurred the diagnosis cost c or not (the profit made on a blind recommendation is p, the profit made on a serious diagnosis is $p - c$), on whether the consumer accepted to be treated or not (if he accepted, then the expert gets in addition to the profit made on the diagnosis the difference between treatment price and treatment cost), and on whether the treatment provided was sufficient to solve the problem or not (if it was not, and if the expert's tariff stipulates a transfer in case of failure, this transfer payment reduces the profit made on the recommendation and on the treatment provided). By assumption, an expert recommends the appropriate treatment if she is indifferent between recommending the appropriate and recommending the wrong treatment, and this fact is common knowledge among all market participants.[4]

Throughout the paper we restrict attention to situations where the following two conditions hold

$$\begin{aligned} v &> \overline{c} + s \\ \overline{c} - \underline{c} &\geq s \end{aligned}$$

The first of these inequalities says that it is efficient to treat both types of problem even in period two and the second inequality is to rule out uninteresting cases. Without this last restriction consumers will never visit more

[4]Introducing some guilt disutility associated with recommending the wrong treatment would yield the same qualitative results as this common knowledge assumption provided the effect is small enough to not outweigh the pecuniary incentives.

than one treatment provider. Throughout the paper we also assume that treatment providers cannot commit to provide treatments at prices below cost. This means that we can restrict attention to prices satisfying the following two conditions

$$\underline{p}, \underline{q} \geq \underline{c}$$
$$\overline{p}, \overline{q} \geq \overline{c}.$$

To keep the analysis simple we finally assume that experts cannot charge a negative diagnosis fee[5]

$$p \geq 0$$

We begin our analysis with the above basic model with homogeneous consumers. Later (in Section 5) we explore the consequences of consumer heterogeneity. There, we first assume that consumers differ in their search cost s. The unit cost of search is assumed to be distributed according to some cumulative distribution function $F(.)$ on some interval $[\underline{s}, \overline{s}]$. Consumers know their search cost, treatment providers know only the distribution.[6] In a second modification of the basic model consumers are assumed to differ in their probability h of needing the expensive treatment. The probabilities h are assumed to be drawn independently from the same cumulative distribution function $G(\cdot)$, with differentiable strictly positive density $g(\cdot)$ on $[0, 1]$. Again, $G(\cdot)$ is assumed to be common knowledge, but a consumer's h is the consumer's private information.[7]

[5] If experts charge a negative diagnosis fee, consumers might have an incentive to engage in 'diagnosis shopping'. To remove this incentive p must exceed $-s$. Our stronger assumption $p \geq 0$ simplifies the analysis but is not important for our main findings.

[6] Car owners know their opportunity cost of searching for and visiting a garage, auto mechanics know only the distribution. Similarly, buyers of complex goods know their remorse on switching seller after receiving a recommendation, expert sellers know only the distribution.

[7] Car owners know how they treat their vehicles and the associated risk of needing certain repairs, auto mechanics know only the distribution. Similarly, buyers of PCs know their profession and the associated 'risk' of needing certain features, expert sellers know only the distribution.

The equilibrium concept we employ is that of perfect Bayesian equilibrium. That is, we require that the strategies of the market participants yield a Bayes-Nash equilibrium not only for each proper subgame, but also for continuation games that are not proper subgames (because they do not stem from a singleton information set).[8] Our focus will be on symmetric equilibria.

Throughout our analysis we use the following notation We use the term Δ to denote the mark-up an expert charges on the diagnosis (that is, $\Delta = p - c$). Similarly, we will use the term $\underline{\Delta}$ for the mark-up the expert charges on the minor, and the term $\overline{\Delta}$ for the mark-up she charges on the major treatment (that is, $\underline{\Delta} = \underline{p} - \underline{c}$ and $\overline{\Delta} = \overline{p} - \overline{c}$).

3 A Benchmark Solution

Let us begin with a characterization of the efficient diagnosis and treatment policy. We then proceed by showing that the efficient solution could be sustained in equilibrium if experts' diagnosis effort was observable and verifiable.

Since searching for a service provider is costly, efficiency requires that consumers are treated by the first provider they visit (that is, separation of diagnosis and treatment is inefficient). Thus, three policies are candidates for the efficient solution

1. performing a serious diagnosis and providing the diagnosed treatment
2. performing no diagnosis and ("blind") provision of $\overline{c}$
3. performing no diagnosis and ("blind") provision of $\underline{c}$; if the treatment fails, then performing no diagnosis and ("blind") provision of $\overline{c}$.

Which of these three policies is the most efficient one? To answer this question we introduce the concept of *generalized cost.* Generalized cost is

[8] Here note that a consumer who visits an expert has to decide whether to stay or to leave without knowing whether the better-informed expert has recommended the right or the wrong treatment.

the expected cost of following the policy under consideration. It includes the search cost, the cost of the treatment(s) provided, the diagnosis cost c if diagnosis is performed, and the expected loss due to treatment failure. Thus, the generalized cost for diagnosis plus efficient treatment is $s+c+(1-h)\underline{c}+h\overline{c}$, the generalized cost for blind provision of $\overline{c}$ is $s+\overline{c}$, and the generalized cost for blind provision of $\underline{c}$ is $s+\underline{c}+h(v+\overline{c})$.[9] The most efficient policy is the policy that minimizes generalized cost. Thus, policy

1. is efficient iff $c \leq \min\{(1-h)(\overline{c}-\underline{c}); h(v+\underline{c})\}$.

2. is efficient iff $(1-h)(\overline{c}-\underline{c}) \leq \min\{h(v+\underline{c}); c\}$.

3. is efficient iff $h(v+\underline{c}) \leq \min\{(1-h)(\overline{c}-\underline{c}); c\}$.

Figure 1 displays the efficient policy for different (c,h) combinations, holding v, $\overline{c}$, $\underline{c}$ and s fixed. Below the $h = 1-c/(\overline{c}-\underline{c})$ line, serious diagnosis and efficient treatment (Policy 1) is more efficient than blind provision of $\overline{c}$ (Policy 2). Below the $h = c/(v+\underline{c})$ line, blind provision of $\underline{c}$ (Policy 3) is more efficient than serious diagnosis and efficient treatment (Policy 1). And below the $h = (\overline{c}-\underline{c})/(v+\overline{c})$ line, blind provision of $\underline{c}$ (Policy 3) is more efficient than blind provision of $\overline{c}$ (Policy 2). Thus, Policy 1 is optimal in Region A, Policy 2 is optimal in Region B, and Policy 3 is optimal in Region C. We will refer back to this figure when discussing the equilibria of our model with unobservable diagnosis effort.

Before turning to this model we first show that the efficient solution could be sustained in equilibrium if experts' diagnosis effort was observable and verifiable. We record this result as

[9] Here notice that we assume that a consumer does not incur another search cost if he buys $\overline{c}$ after first having tried $\underline{c}$. In an earlier version of this paper (Dulleck and Kerschbamer 2005b) we employed the alternative assumption that visiting a provider always costs s. The analysis is slightly more complicated, the qualitative results are the same, however.

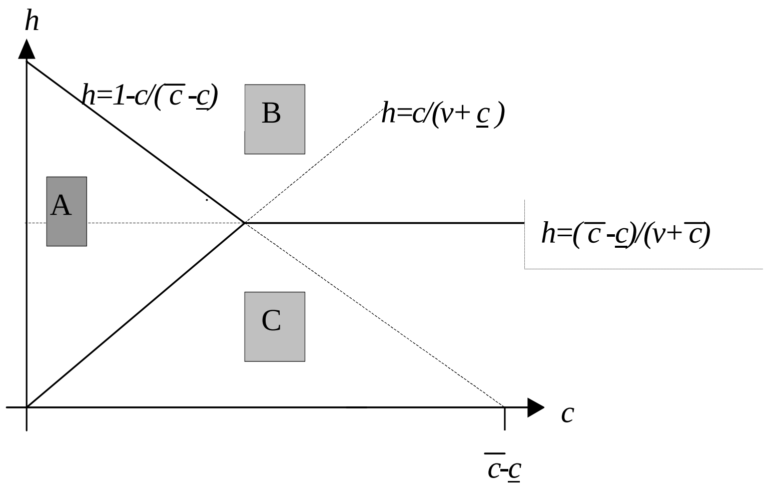

Figure 1: Efficient Policy Policy 1 in Region A, Policy 2 in Region B, and Policy 3 in Region C

Proposition 1 *If experts' diagnosis effort is observable and verifiable then in any equilibrium the market will be efficient. In one equilibrium experts and discounters charge marginal cost prices for diagnosis and treatment.*

Proof. Obvious from the discussion below and therefore omitted. ■

The intuition behind the efficiency result of Proposition 1 is easily provided Consumers who visit a discounter face no incentive problem. Everything is as if discounters just provided normal goods. Thus, if the parameters of the model are such that we are in Region C, then in any equilibrium $\underline{q} = \underline{c}$ and $\overline{q} = \overline{c}$ by the usual price-undercutting argument. Similarly, if the parameters of the model are such that we are in Region B, then in any equilibrium $\overline{q} = \overline{c}$ by the usual price-undercutting argument. With these prices experts cannot attract consumers in the B or C region without making losses.[10] There

[10]The only way for experts to attract customers without making losses in this situation is to act like a discounter; that is, to offer treatments at marginal cost, without providing a serious diagnosis. Here note that although in our model experts and discounters are assumed to be distinct providers, nothing would change if we assumed instead that there is only one kind of treatment provider with the characteristics we have ascribed to experts

remain consumers in Region A. In an efficient solution they should visit an expert, the expert should make a serious diagnosis, the expert should then recommend the diagnosed treatment and consumers should decide to receive it. Under the conditions of Proposition 1 experts' diagnosis effort is contractible, inducing the expert to provide a serious diagnosis is therefore no problem. There remains (i) experts' incentive to recommend the wrong treatment, and (ii) consumers' incentive to reject the treatment recommended by an expert and to visit another provider. In equilibrium incentive (i) is removed by experts posting prices and transfers satisfying $(\underline{\Delta} - \overline{\Delta}) \in [0, t]$; and incentive (ii) is removed by experts committing to prices yielding $\underline{\Delta}$, $\overline{\Delta} \in [0, s]$. The latter result is trivial given discounters' incentive to attract consumers who know what they need. The intuition for the former result is as follows If an expert posts prices violating $(\underline{\Delta}-\overline{\Delta}) \in [0, t]$, consumers would become suspicious; they would correctly infer that the expert will either always recommend the major treatment (if $\underline{\Delta} < \overline{\Delta}$), or always recommend the minor one (if $\underline{\Delta} - t > \overline{\Delta}$), and they would adjust their willingness to pay accordingly. So, experts cannot gain from cheating. Consequently, they post prices that induce non-fraudulent behavior. With prices that induce non-fraudulent behavior we are again back to the normal good case; that is, Bertrand competition yields prices such that underbidding yields losses and charging more implies a loss of customers. Putting these conditions together yields prices p, $\underline{p}$ and $\overline{p}$ and transfers t fulfilling the following properties $\Delta + (1-h)\underline{\Delta} + h\overline{\Delta} = 0$; $(\underline{\Delta} - \overline{\Delta}) \in [0, t]$; $\underline{\Delta}$, $\overline{\Delta} \in [0, s]$; and $\Delta \geq -c$. Obviously, a tariff with $\Delta = \underline{\Delta} = \overline{\Delta} = t = 0$ satisfies all these conditions. Given that each of the three policies is available in equilibrium at marginal cost and that inefficient policies (such as diagnosis shopping or separation of diagnosis and treatment) are unattractive for consumers, consumers will choose the efficient policy.

and if we call such a provider 'expert' if she sets either $p > 0$ or $t > 0$, and 'discounter' otherwise. All results remain unaffected provided that at least four of these treatment providers populate the market. In what follows we call an expert who acts like a discounter a discounter.

The efficiency result of Proposition 1 is not sensitive on whether customers are homo- or heterogeneous. If consumers differ in their search cost and if the parameters of the model are such that we are in Region A, then prices yielding $\Delta = \underline{\Delta} = \overline{\Delta} = 0$ are the unique equilibrium strategies for experts if at least one customer has $s = 0$. Why? Because any cross-subsidization of diagnosis with treatment prices (that is, $\Delta < 0$ and either $\underline{\Delta} > 0$ or $\overline{\Delta} > 0$) would induce low-search-cost consumers to demand a diagnosis by an expert and then to buy the treatment by a discounter (see the discussion in Subsection 5.1). If consumers differ in their probabilities of needing different treatments then any deviation from $\Delta = \underline{\Delta} = \overline{\Delta} = 0$ might invite consumers outside Region A to inefficiently consult an expert. Why? Again, because any deviation from $\Delta = \underline{\Delta} = \overline{\Delta} = 0$ (holding $\Delta + h\overline{\Delta} + (1-h)\underline{\Delta}$ constant) implies some kind of cross subsidization. This cross-subsidization invites some consumers who should efficiently buy at a discounter to consult an expert for diagnosis (see the discussion in Subsection 5.2).

4 The Homogeneous Customers Case: Experts or Discounters

We now turn to our basic model with unobservable diagnosis effort. We begin with the homogeneous consumers case. Obviously, if the parameters of the model are such that we are either in Region B or in Region C of Figure 1, then the equilibrium behavior of market participants does not depend on whether the experts' diagnosis effort is observable or not. In both cases only discounters are active and they charge marginal cost prices. Our main focus in the rest of the paper will therefore be on parameter constellations in Region A. An important question in this region is whether experts can signal their diagnosis effort through their choice of diagnosis and treatment prices as well as of transfer payments (for the case of treatment failure). The answer turns out to be yes, but at the cost of being vulnerable to competition by discounters. To see this, first observe that the most attractive options for an expert who gets visited by a consumer and who expects to be able to induce the consumer to accept the treatment she recommends are now to a) seriously diagnose the customer and recommend the appropriate treatment, b) not diagnose the customer and (blindly) recommend $\overline{c}$, and c) not diagnose the customer and (blindly) recommend $\underline{c}$. Given that the expert is free to set the transfer, she will use it to signal that option c) is unattractive for her. For option a) to dominate option c) the transfer payment t and the mark-ups Δ, $\underline{\Delta}$ and $\overline{\Delta}$ need to fulfill the condition $\Delta + (1-h)\underline{\Delta} + h\overline{\Delta} \geq p + \underline{\Delta} - ht$ which is equivalent to

$$t \geq \underline{\Delta} - \overline{\Delta} + \frac{c}{h}. \tag{1}$$

This condition can always easily be met. That is, an expert can always easily signal with her choice of the warranty payment t that she has no incentive to choose option c). Option b) is the more critical one. For option a) to dominate option b) the mark-ups Δ, $\underline{\Delta}$ and $\overline{\Delta}$ need to fulfil the condition $\Delta + (1-h)\underline{\Delta} + h\overline{\Delta} \geq p + \overline{\Delta}$ which is equivalent to

$$\frac{c}{1-h} \leq \underline{\Delta} - \overline{\Delta}. \tag{2}$$

In words The mark-up on the minor intervention must exceed the mark-up on the major one by such an amount that the expert can earn the diagnosis cost on selling the minor intervention. If experts were able to commit to provide treatments at prices below cost no problem would arise. But given experts' commitment problem, prices need in addition fulfill $\underline{\Delta}, \overline{\Delta} \geq 0$, which, together with condition (2), yields

$$\frac{c}{1-h} \leq \underline{\Delta}. \tag{3}$$

Now consider consumers. They are aware that discounters charge marginal cost prices. Consequently, they will accept to receive the treatment recommended by an expert only if the price the expert charges for the recommended treatment does not exceed the sum of treatment cost plus search cost. This implies another restriction on the price for the minor intervention, namely

$$\underline{\Delta} \leq s. \tag{4}$$

Obviously, if $s < \frac{c}{1-h}$ then conditions (3) and (4) are incompatible. This leads us to our next result

Proposition 2 *Consider our basic model with homogeneous consumers and unobservable diagnosis effort. Suppose that the parameters of the model are such that we are in Region A of Figure 1. Then the efficient solution is sustainable in equilibrium if and only if* $s \geq \frac{c}{1-h}$. *If* $s < \frac{c}{1-h}$ *then experts refrain from providing advice and the market is served by discounters.*

Proof. From the discussion above it is clear that experts cannot survive as full service providers (i.e., diagnosis and treatment providers) whenever $s < \frac{c}{1-h}$. For $s \geq \frac{c}{1-h}$ prices and transfers satisfying conditions (1) and (2) above, as well as $p \geq 0, \overline{\Delta} \geq 0$ and $\Delta + (1-h)\underline{\Delta} + h\overline{\Delta} = 0$ are the unique equilibrium prices of experts by the usual price-undercutting argument. These conditions together yield $p = 0$, $\underline{p} = \underline{c} + \frac{c}{1-h}$, $\overline{p} = \overline{c}$ and $t \geq \frac{c}{(1-h)h}$. ■

Let us recapitulate the intuition for the inefficiency result of Proposition 2. With unobservable diagnosis effort experts must be prevented from choosing one of the following two cheating strategies abstaining from diagnosis and potentially undertreating the consumer (that is, blindly recommending $\underline{c}$), and abstaining from diagnosis and potentially overtreating the consumer (that is, blindly recommending $\overline{c}$). The undertreating incentive is easily removed by experts offering a warranty for the case of treatment failure. To remove the overtreating incentive, charging the diagnosis fee without providing one and selling the major treatment for sure must be unprofitable. This is only possible if the mark-up on the major intervention is set to zero and if the diagnosis is given for free. This in turn implies that the diagnosis cost needs to be carried by $\underline{p}$. If this is impossible because the necessary mark-up is below the "switching" cost of consumers, then experts cannot survive as full service providers on the market.

How does the new equilibrium look like? Figure 2 provides the answer. As compared to Figure 1, the original Region A is split into three distinct parts. If consumers' switching cost is above the mark-up on the minor intervention

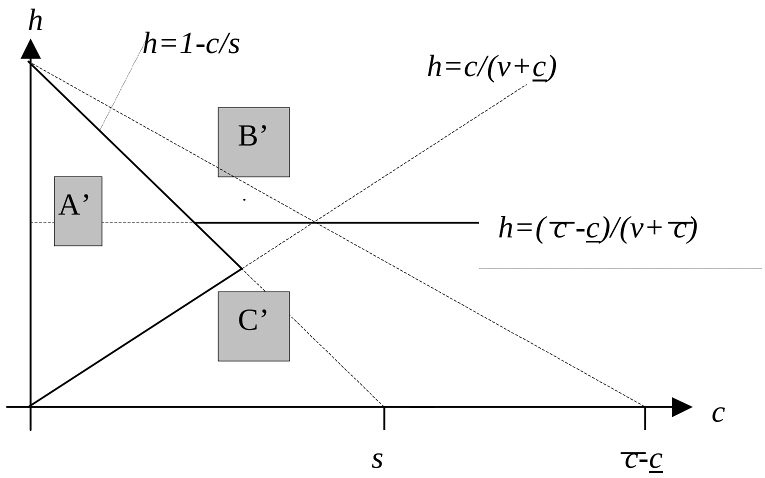

Figure 2: Market Equilibria with Homogeneous Consumers and Unobservable Diagnosis Effort

necessary to finance the diagnosis cost, then experts provide serious diagnosis and appropriate treatment and full efficiency prevails (Region A'). Otherwise, inefficiencies arise. In area $A \cap B'$ experts should but do not provide diagnosis and customers blindly buy $\bar{c}$ from a discounter. As compared to the first best policy, this leads to an efficiency loss of $(1-h)(\bar{c}-\underline{c})-c$. Similarly, in area $A \cap C'$ experts should but do not provide diagnosis and consumers blindly buy $\underline{c}$ from a discounter. As compared to the first best policy, this implies an efficiency loss of $h(v+\underline{c})-c$.

5 Heterogenous Customers and Market Unraveling

Up to now our focus was on the basic model with homogeneous consumers. In this section we show that the inefficiencies of the homogeneous consumers case amplify if consumers are heterogeneous. We begin with a scenario in which consumers differ in their search cost. Later we relax the assumption that consumers have the same expected cost of efficient treatment.

5.1 Heterogeneity in the Search Cost: Who Free Rides on Experts' Advice?

In this subsection we study a scenario in which consumers differ in their search cost s. The unit cost of search is assumed to be distributed according to a continuous distribution function $F(.)$ on some interval $[\underline{s}, \overline{s}]$, with $0 \leq \underline{s} < \overline{s} \leq \overline{c} - \underline{c}$. Consumers know their search cost, treatment providers know only the distribution.

The main issue with consumers differing in their search cost is a domino effect triggered by low-search-cost consumers' free-riding on experts' advice. This domino effect might lead to a complete unraveling of the experts' market. To see the problem, consider a market in which all consumers are located in Region A of Figure 1 and in which some of the consumers have low ($s < \frac{c}{1-h}$), others high search cost ($s > \frac{c}{1-h}$). Then low search cost consumers have an incentive to consult an expert to get a free diagnosis and to switch to a discounter if the expert recommends the minor treatment. But if some customers switch after receiving the free diagnosis then experts must increase the mark-up on the minor intervention to finance the free-riding customers' diagnosis effort. This price-increase might lead even more customers to free ride on the diagnosis effort. As the discrete example below reveals the resulting domino effect might lead to a complete unraveling of the experts' market. In the example, there are one hundred consumers with parameter constellations in Region A. That is, the entire market should efficiently be served by an expert. Ninety-nine consumers are located in Region A' (i.e., for 99 consumers $s \geq \frac{c}{1-h}$), one is located in area $A \cap B'$ (i.e., for 1 consumer $s < \frac{c}{1-h}$). If only the 99 consumers located in Region A' were on the market, full efficiency would prevail. That is, each consumer would get serious diagnosis and appropriate treatment. After adding the consumer located in area $A \cap B'$, the market unravels completely such that no expert survives competition by discounters.

Example 1 *Each consumer has equal chances of having the minor and having the major problem* ($h = 0.5$)*. The cost of treating the minor problem is zero* ($\underline{c} = 0$)*, the cost of treating the major problem is six* ($\overline{c} = 6$)*. Consumers' valuation for a successful intervention is ten* ($v = 10$)*, the diagnosis cost is one* ($c = 1$)*. Consumers differ in their search cost s. First suppose that there are 99 consumers in the market, with search costs* 2.00, 2.01, 2.02, ..., 2.96, 2.97, 2.98. *Then experts post prices and transfers satisfying $p = 0$, $\underline{p} = \underline{c} + \frac{c}{1-h} = 2$, $\overline{p} = \overline{c} = 6$ and $t \geq \frac{c}{(1-h)h} = 4$ and efficiently serve their customers. Now let an additional consumer with search cost $s = 1.99$ appear on the market. If experts anticipate that this consumer will free-ride on their diagnosis effort, they will increase the price of the minor intervention to $\underline{p} = \underline{c} + \frac{c}{.99(1-h)} = 2.0202$ to avoid losses. But at this price there is not one free-riding consumer, there are four of them. With four free-riding consumers the price for the minor intervention has to be at least $\underline{p} = \underline{c} + \frac{c}{.96(1-h)} = 2.0833$ to avoid losses. But then there are ten free riding consumers, implying a cost-covering price for the minor intervention of at least $\underline{p} = \underline{c} + \frac{c}{.90(1-h)} = 2.2222$. With this price there are twenty-four free-riders implying that the price for the minor intervention must increase to $\underline{p} = \underline{c} + \frac{c}{.76(1-h)} = 2.6315$ to avoid losses. But at this price 65% of the consumers are free-riders so that the price for the minor intervention must exceed $\underline{p} = \underline{c} + \frac{c}{.35(1-h)} = 5.7143$ to cover the diagnosis cost. At this price no consumer will ever accept a recommendation for the minor intervention.*

Is the domino effect cropping up in Example 1 a robust phenomenon or purely a pathology of the numbers used in the example? Our next two results (Proposition 3 and Implication 1) help to answer this question. Both results look into cases where (i) all consumers are located in Region A of Figure 1 and where (ii) $\underline{s} \leq \frac{c}{1-h} \leq \overline{s}$. Condition (i) implies that in the efficient solution all consumers are served by an expert; and condition (ii) ensures that in an equilibrium of our model with unobservable diagnosis effort, all consumers with a search cost $s \in [\frac{c}{1-h}, \overline{s}]$ would still be served by an expert if s was observable and experts could reject to treat customers with a too low s.[11] The first result (Proposition 3) looks at an instance where some consumers are located in Region A' of Figure 2 (i.e., for some consumers $s \geq \frac{c}{1-h}$), while some other consumers are located either in area $A \cap B'$ or in area $A \cap C'$ (i.e., for some consumers $s < \frac{c}{1-h}$).

Proposition 3 *Consider our model with unobservable diagnosis effort. Suppose that consumers differ in their search cost s and that the unit cost of search is distributed according to some continuous distribution function $F(\cdot)$, with differentiable strictly positive density $f(\cdot)$ on $[\underline{s}, \overline{s}]$, with $0 \leq \underline{s} < \overline{s} \leq \overline{c} - \underline{c}$. Further suppose that the parameters of the model are such that all consumers are located in Region A of Figure 1 and that $\frac{c}{1-h} \in [\underline{s}, \overline{s}]$. Then experts can survive as full service providers on the market if and only if* $\max(s(1 - F(s)) \geq \frac{c}{1-h}$.

Proof. In an (interior) solution, in which experts are active as full service providers on the market, all consumers consult an expert for diagnosis, some accept a recommendation for the minor intervention and some others reject such a recommendation. Hence, by the continuity of $F(.)$, if such an interior solution exists, then there exists some critical consumer $\widetilde{s} \in [\underline{s}, \overline{s}]$ who is exactly indifferent between staying with the expert and buying from a discounter when being told that the minor intervention is sufficient to solve his problem. This critical consumer is given by $\widetilde{s} = \underline{\Delta}$. Furthermore, if an interior solution exists, then experts make non-negative (experts are free to post prices that attract no consumers) and non-positive (by the usual price-undercutting argument) profits in it. Thus, since in any equilibrium in which

[11] Obviously, if all consumers were located in the same region of Figure 2 then heterogeneity of consumers wouldn't change anything in the equilibrium behavior of market participants. That is, if $\underline{s} \geq \frac{c}{1-h}$ and $h \geq \frac{c}{v+\underline{c}}$ (i.e., if all consumers are located in Region A') then all consumers would efficiently be served by an expert. And if either $\overline{s} < \frac{c}{1-h}$ or $h < \frac{c}{v+\underline{c}}$ (i.e., if all consumers are either located in Region B' or in Region C') then only discounters (or experts who act like discounters) would be active in the market.

experts act as full service providers on the market $p = 0$ and $\overline{\Delta} = 0$, the equilibrium mark-up on the minor intervention must be equal to $\frac{c}{(1-h)(1-F(\widetilde{s}))}$. Putting these two equations together yields $\widetilde{s} = \frac{c}{(1-h)(1-F(\widetilde{s}))}$. Given that $s(1 - F(s))$ is equal to $\underline{s}$ for $s = \underline{s}$ and equal to 0 for $s = \overline{s}$, and given that $0 \leq \underline{s} \leq \frac{c}{1-h}$, no interior solution exists iff $\max(s(1 - F(s)) < \frac{c}{1-h}$.[12] ■

Our next result (Implication 1) looks at an instance where (as in Example 1), starting from a situation where all consumers are located in Region A' of Figure 2, a minimal decrease in the search cost of the marginal consumer (or a minimal increase in the diagnosis cost) leads to a discrete switch from an (efficient) "experts only" to an (inefficient) "discounter only" market. In the result, reference is made to the hazard rate of the distribution $F(.)$. The hazard rate is defined by $h(s) = \frac{f(s)}{1-F(s)}$.

Implication 1 *Suppose that the general conditions of Proposition 3 hold. Further suppose that the hazard rate of the distribution $h(s)$ satisfies $h(s) > \frac{1}{s}$ for all $s \in [\underline{s}, \overline{s}]$. Then a marginal ($\varepsilon$) decrease of $\underline{s}$ from $\underline{s} = \frac{c}{(1-h)}$ to $\underline{s} = \frac{c}{(1-h)} - \varepsilon$ leads to a discrete switch from an (efficient) "experts only" to an (inefficient) "discounter only" market.*

Proof. For $\underline{s} = \frac{c}{(1-h)}$ all consumers are located in Region A' of Figure 2. Thus, prices and transfers satisfying conditions (1) and (2), as well as $p \geq 0$, $\overline{\Delta} \geq 0$ and $\Delta + (1 - h)\underline{\Delta} + h\overline{\Delta} = 0$ are the unique equilibrium prices of experts by the usual price-undercutting argument. These conditions together yield $p = 0$, $\underline{p} = \underline{c} + \frac{c}{1-h}$, $\overline{p} = \overline{c}$ and $t \geq \frac{c}{(1-h)h}$. With these prices all consumers consult an expert and buy the appropriate treatment there. Thus, full efficiency prevails. That efficient equilibria cease to exist whenever $\underline{s} < \frac{c}{(1-h)}$ follows from $\underline{s}(1 - F(\underline{s})) = \underline{s} < \frac{c}{1-h}$ and from the fact that $s(1 - F(s))$ is monotonically decreasing whenever $h(s) > \frac{1}{s}$. Thus, $\max(s(1 - F(s)) = \underline{s} < \frac{c}{(1-h)}$. ■

It is not difficult to find distribution functions which have the property asked for in Implication 1. The uniform distribution, for instance, exhibits

[12] Here note that the same argument also implies that if interior solutions exist, their number must be even.

this property whenever $2\underline{s} > \overline{s}$

Example 2 *Consider the framework of Example 1. In contrast with Example 1, assume that consumers' search cost s is uniformly distributed on* $[2, 3]$*. Then the market is efficiently served by experts who post prices and transfers satisfying* $p = 0$, $\underline{p} = \underline{c} + \frac{c}{1-h} = 2$, $\overline{p} = \overline{c} = 6$ *and* $t \geq \frac{c}{(1-h)h} = 4$*. Furthermore, a marginal decrease in $\underline{s}$ (or a marginal increase in c) leads to a discrete switch from an (efficient) "experts only" to an (inefficient) "discounter only" market.*

5.2 Heterogeneity in the Expected Cost of Efficient Treatment: Who Visits an Expert?

The inefficiencies of the homogeneous consumers case are also amplified if consumers differ in their expected cost of efficient treatment. To show this, we study a model where consumers differ in their probability h of needing the expensive treatment. The probabilities h are assumed to be drawn independently from the same continuous distribution function $G(\cdot)$, with differentiable strictly positive density $g(\cdot)$ on $[0, 1]$. Again, $G(\cdot)$ is assumed to be common knowledge, but a consumer's h is the consumer's private information.

The main topic with consumers differing in their expected cost of efficient treatment is that the wrong segment of the market visits an expert. To separate this adverse selection problem from the free-rider problem studied in the previous subsection, we assume in this subsection that consumers' search cost s is large enough to avoid customer switching. A sufficient condition for this to be the case is $s \geq \overline{c} - \underline{c}$. Under this condition there is no difference between Figure 1 and Figure 2. That is, with homogeneous customers, full efficiency would prevail. As we will see below, this is no longer the case if consumers differ in their expected cost of efficient treatment.

Obviously, no new problems would arise, if all consumers were located either in Region B or in Region C of Figure 1. Then only discounters (or experts who act like discounters) would offer treatments at prices equal to marginal cost and full efficiency would prevail. We therefore assume in this subsection that the diagnosis cost c satisfies $0 < c < \frac{(\overline{c}-\underline{c})(v+\underline{c})}{(v+\overline{c})}$. Looking at Figure 1, this assumption assures that all consumers are threaded up on a vertical line to the left of the intersection of the three straight lines separating regions A, B and C. That is, in an efficient solution, consumers with an h in $[0, c/(v+\underline{c})]$ should blindly buy $\underline{c}$ from a discounter, consumers in $[c/(v+\underline{c}), 1-c/(\overline{c}-\underline{c})]$ should get serious diagnosis and appropriate treatment from an expert, and consumers in $[1-c/(\overline{c}-\underline{c}), 1]$ should blindly buy $\overline{c}$ from a discounter.

Is the efficient solution sustainable in equilibrium of our model with unobservable diagnosis effort? The answer turns out to be no. The reason is, that in order to signal that they have the right incentive to perform serious diagnosis and to provide appropriate treatment, experts have to charge prices that imply a cross-subsidization of high h consumers by low h consumers. This cross-subsidization invites (all) consumers who should blindly buy $\overline{c}$ from a discounter to consult an expert for diagnosis and (some) consumers who should visit an expert for diagnosis to blindly buy $\underline{c}$ from a discounter. Anticipating this adverse selection problem, experts increase the mark-up on the minor intervention to avoid losses. This price-increase might again set in motion a chain-reaction similar to the one studied in the previous subsection. As the discrete example below reveals, this chain reaction can again lead to a complete unraveling of the experts' market.

Example 3 *The cost of the minor treatment is zero ($\underline{c} = 0$), the cost of the major treatment is twenty ($\overline{c} = 20$), the diagnosis cost is seven ($c = 7$). Each consumer has a valuation for a successful intervention of forty ($v = 40$) and a search cost of twenty ($s = 20$). Consumers differ in their probability h of needing the major treatment. Each consumer's h takes with equal probability one of the following twenty values* 0.025, 0.075, 0.125, 0.175, ..., 0.825, 0.875, 0.925, 0.975. *In an efficient solution, consumers in* [0.025, 0.125] *should blindly buy $\underline{c}$ from a discounter, consumers in* [0.175, 0.625] *should get* *serious diagnosis and appropriate treatment from an expert, and consumers in* $[0.675,\ 0.975]$ *should blindly buy* $\overline{c}$ *from a discounter. First suppose that experts naively focus on that segment of the market that efficiently should be served by experts. The average* h *in this segment is* 0.4*. So, the mark-up on the minor intervention must be at least* 11.667 *to cover the diagnosis cost. But, with prices satisfying* $p = 0$, $\underline{p} = 11.667$ *and* $\overline{p} = \overline{c} = 20$ *consumers in* $[0.175,\ 0.225]$ *inefficiently decide for blind provision of* $\underline{c}$ *and consumers in* $[0.675,\ 0.975]$ *inefficiently decide for a visit by an expert. Thus, with these prices the average* h *of consumers visiting an expert is not* 0.4 *but rather* 0.625*. But then the price for the minor intervention must exceed* $\underline{p} = 18.667$ *to avoid losses. At this price another low* h *consumer (the one with* $h = 0.275$*) also decides against a visit by an expert implying a cost-covering price for the minor intervention of* $\underline{p} = 20$*. At this price another low* h *consumers decides against a visit by an expert, so* $\underline{p}$ *should increase again to avoid losses. But with prices satisfying* $p = 0$, $\underline{p} > \overline{c}$ *and* $\overline{p} = \overline{c} = 20$ *no consumer will ever visit an expert.*

What can we learn from this discrete example for the continuous case? First, if there is an equilibrium in which experts act as full service providers in the market, then no consumer will ever blindly buy $\overline{c}$ from a discounter. Why? Because experts give the diagnosis for free and charge marginal cost prices for $\overline{c}$, and because in any equilibrium in which experts act as full service providers in the market we have $\underline{p} \leq \overline{p}$. So, the worst case for a consumer who visits an expert is that he has to pay the same amount he would have paid for $\overline{c}$ at the discounter in any case. Secondly, in any equilibrium, some consumers will always blindly buy $\underline{c}$ from a discounter. Why? Because $\underline{p} > \underline{c}$ and $\underline{q} = \underline{c}$, and because for some consumers it is extremely unlikely that they need the major intervention. For the continuous case, these two facts together imply that in any equilibrium in which experts act as full service providers on the market, there must exist some critical consumer $\widetilde{h} \in (0,1)$ such that all consumers in $[\widetilde{h}, 1]$ visit an expert while all consumers in $[0, \widetilde{h})$

blindly buy $\underline{c}$ from a discounter. This critical consumer is given by

$$\widetilde{h} = \frac{\underline{\Delta}}{v + \underline{p}}. \tag{5}$$

Also, if there is an equilibrium in which experts act as full service providers on the market, then they make non-positive and non-negative profits in it. Thus, since in any equilibrium in which experts act as full service providers on the market $p = 0$ and $\overline{\Delta} = 0$, the equilibrium mark-up on the minor intervention must be equal to

$$\underline{\Delta} = \frac{c}{1 - E(h \left| h \geq \widetilde{h} \right)} \tag{6}$$

Putting equations (5) and (6) together yields

$$\frac{\widetilde{h}[1 - E(h \left| h \geq \widetilde{h} \right)]}{(1 - \widetilde{h})} = \frac{c}{(v + \underline{c})}. \tag{7}$$

Equation (7) yields a value for $\widetilde{h}$ and (together with equation (5)) a value for $\underline{p}$.[13] For these values to be feasible as a solution, they must satisfy $\widetilde{h} \in [0, 1]$ and $\underline{p} \leq \overline{p} = \overline{c}$. By equation (5) $\underline{p} \leq \overline{c}$ whenever $\widetilde{h} \leq \frac{\overline{c} - \underline{c}}{v + \overline{c}}$. Since $\overline{c} - \underline{c} < v + \overline{c}$ this condition is more demanding than $\widetilde{h} \leq 1$. It follows, that experts can survive as full service providers on the market if and only if there exists an $\widetilde{h} \leq \frac{\overline{c} - \underline{c}}{v + \overline{c}}$ such that $\frac{\widetilde{h}[1 - E(h \left| h > \widetilde{h} \right)]}{(1 - \widetilde{h})} = \frac{c}{(v + \underline{c})}$. We record this as

Proposition 4 *Consider our model with unobservable diagnosis effort. Suppose that consumers differ in their probability h of needing the expensive treatment and that h is distributed according to some continuous distribution function $G(\cdot)$, with differentiable strictly positive density $g(\cdot)$ on $[0, 1]$. Further suppose that the parameters of the model are such that $0 < c < \frac{(\overline{c} - \underline{c})(v + \underline{c})}{(v + \overline{c})}$ and $s \geq \overline{c} - \underline{c}$. Then experts can survive as full service providers on the market if and only if there exists an $\widetilde{h} \leq \frac{\overline{c} - \underline{c}}{v + \overline{c}}$ such that $\frac{\widetilde{h}[1 - E(h \left| h \geq \widetilde{h} \right)]}{(1 - \widetilde{h})} = \frac{c}{(v + \underline{c})}$.*

[13] That Equation (7) yields at least one solution for $\widetilde{h}$ is shown in the Appendix.

Proof. If there exists an $\widetilde{h} \leq \frac{\overline{c}-\underline{c}}{v+\overline{c}}$ such that $\frac{\widetilde{h}[1-E(h\left|h\geq\widetilde{h}\right)]}{(1-\widetilde{h})} = \frac{c}{(v+\underline{c})}$ then experts can post prices and transfers satisfying $p = 0$, $\underline{\Delta} = \frac{c}{1-E(h\left|h\geq\widetilde{h}\right)}$, $\overline{\Delta} = 0$ and $t \geq \frac{c}{(1-h)h}$. With these prices, consumers in $[\widetilde{h}, 1]$ will visit an expert and get serious diagnosis and appropriate treatment there, and consumers in $[0, \widetilde{h})$ will blindly buy $\underline{c}$ from a discounter. With the market segment $[\widetilde{h}, 1]$ as customers, experts will on average make non-positive and non-negative profits as required in equilibrium. If there exists no $\widetilde{h} \leq \frac{\overline{c}-\underline{c}}{v+\overline{c}}$ such that $\frac{\widetilde{h}[1-E(h\left|h\geq\widetilde{h}\right)]}{(1-\widetilde{h})} = \frac{c}{(v+\underline{c})}$ then experts cannot survive as full service providers on the market by the arguments given above. ■

The condition given in Proposition 4 for the existence of an equilibrium in which experts survive as full service providers on the market is not very transparent. How does this condition translate to the case where the hs are uniformly distributed on $[0, 1]$? Implication 2 provides an answer

Implication 2 *Suppose that the general conditions of Proposition 4 hold. Further suppose that h is uniformly distributed on $[0, 1]$. Then experts can survive as full service providers on the market if and only if $c \leq (\overline{c} - \underline{c})(v + \underline{c})/2(v + \overline{c})$.*

Proof. For the uniform distribution $E(h\left|h \geq \widetilde{h}\right) = \frac{1+\widetilde{h}}{2}$. Inserting this expression in equation (7) and solving for $\widetilde{h}$ yields $\widetilde{h} = \frac{2c}{v+\underline{c}}$ and (together with equation (5)) $\underline{p} = \underline{c} + \frac{2c(v+\underline{c})}{v+\underline{c}-2c}$. For these values for $\widetilde{h}$ and $\underline{p}$ to be feasible as a solution, they must satisfy $\widetilde{h} \in [0, 1]$ and $\underline{p} \leq \overline{p} = \overline{c}$, or equivalently $c \leq (v + \underline{c})/2$ and $c \leq (\overline{c} - \underline{c})(v + \underline{c})/2(v + \overline{c})$. Since $(\overline{c} - \underline{c}) < (v + \overline{c})$ the second inequality is more demanding than the first one. This completes the argument. ■

6 Vertical Restraints: Overprovision of Diagnosis and Insufficient Treatment

The traditional vertical-restraints-literature typically takes the perspective of a profit-maximizing manufacturer wishing to market its products to consumers through a competitive retail sector.[14] Let us, in this section, take this perspective and ask whether a monopolistic manufacturer – or a cartelized industry – would have incentives and means to correct, or at least ameliorate, the distortions encountered in the previous two sections. To tackle this question we assume that the manufacturer's marginal cost of production for the minor treatment (or the major treatment, respectively) is $\underline{c}$ (or $\overline{c}$, respectively) and that she sells the treatment at wholesale prices $\underline{w}^e$ and $\underline{w}^d$ ($\overline{w}^e$ and $\overline{w}^d$, respectively) to experts and discounters. We interpret the discriminatory pricing on the wholesale level as vertical restraints. For instance, $\underline{w}^d = \overline{w}^d = \infty$ is equivalent to exclusive dealership. We begin with the homogeneous consumers case.

6.1 The Homogeneous Consumers Case

First notice, that with homogeneous consumers, the monopolistic manufacturer has never an incentive to use both, experts and discounters, as distribution channels. Thus, the following policies are natural candidates for a profit maximizing solution

1 Sell both types of treatment, and sell them through experts only ($\underline{w}^d = \overline{w}^d = \infty$); charge wholesale prices $\underline{w}^e$ and $\overline{w}^e$ such that all consumers visit an expert.

2 Sell only the major treatment ($\underline{w}^e = \underline{w}^d = \infty$), and sell it through

[14] An exception is Perry and Besanko (1991) who examine a model with two manufacturers who distribute their products through exclusive retail dealers and who compete for customers indirectly by inducing retailers to carry their product.

discounters only ($\overline{w}^e = \infty$); set the wholesale price $\overline{w}^d$ such that all consumers buy $\overline{c}$ immediately.

3a Sell both types of treatment immediately, and sell them through discounters only ($\underline{w}^e = \overline{w}^e = \infty$); charge wholesale prices $\underline{w}^d$ and $\overline{w}^d$ such that all consumers first try $\underline{c}$, and, if this treatment fails, then buy $\overline{c}$.

What is the maximal profit the manufacturer can earn by employing each of these policies? First remember that Bertrand competition among experts yields $p = 0$, $\underline{p} = \underline{w}^e + c/(1-h)$ and $\overline{p} = \overline{w}^e$. Similarly, Bertrand competition among discounters yields $\underline{q} = \underline{w}^d$ and $\overline{q} = \overline{w}^d$. First consider Policy 1. If a consumer's problem is left untreated, he incurs a cost of $2v$. If he visits an expert in period 1 his cost is $s+(1-h)\ \underline{p}+h\overline{p} = s+c+(1-h)\underline{w}^e+h\overline{w}^e$. Thus, the maximal profit per consumer the manufacturer can earn with Policy 1 is $\pi_1 = 2v - s - c - (1-h)\underline{c} - h\overline{c}$.[15] If the manufacturer employs Policy 2 then she charges $\overline{w}^d = 2v - s$ leading to a profit of $\pi_2 = 2v - s - \overline{c}$. Finally consider Policy 3a. With this policy, prices have to fulfill (i) a period 1 participation constraint ensuring that consumers buy $\underline{c}$ in the first period; (ii) a period 2 participation constraint ensuring that consumers buy $\overline{c}$ in the second period if the low quality treatment failed in the first period; and (iii) a self selection constraint ensuring that customers do not buy $\overline{c}$ in period 1. It is easy to show that (i) is redundant given (ii) and (iii). Thus, since increasing $\overline{w}^d$ relaxes (iii), the manufacturer will set $\overline{w}^d = v$, the maximum value consistent with (ii). With $\overline{w}^d = v$, (iii) yields $\underline{w}^d = (1-2h)v$. Thus, the maximal feasible profit with Policy 3a is $\pi_{3a} = (1-h)v - \underline{c} - h\overline{c}$.[16]

A comparison between π_1, π_2 and π_{3a} reveals that Policy 3a is strictly dominated by Policy 2. The reason is, that the availability of the major

[15] Since consumers' participation constraint has only to hold in expectation, the profit maximizing wholesale prices $\underline{w}^e$ and $\overline{w}^e$ are not uniquely determined. One example for a profit maximizing solution is $\underline{w}^e = 2v - (c+s)/(1-h)$ and $\overline{w}^e = 2v$, leading to consumer prices $\underline{p} = 2v - s/(1-h)$ and $\overline{p} = 2v$.

[16] Remember that we abstract from discounting. Adding a cost to the expert (and the consumer) would add another parameter without qualitatively changing the results.

treatment at a reasonable price in period 1 cannibalizes the market for the minor treatment. Is there a more profitable alternative to Policy 3a? In our simple static framework with a fixed population the following policy is a natural candidate

3b Sell both types of treatment through discounters only ($\underline{w}^e = \overline{w}^e = \infty$), but sell in the first period only $\underline{c}$ and in the second period only $\overline{c}$; charge wholesale prices $\underline{w}^d$ and $\overline{w}^d$ such that all consumers first try $\underline{c}$, and, if this treatment fails, then buy $\overline{c}$.

What is the maximal profit attainable with this policy? Whereas the maximal price the manufacturer can charge for $\overline{c}$ in period 2 remains the same as with Policy 3a ($\overline{w}^d_{t=2} = v$), the self selection constraint becomes redundant because consumers now have to forgo v if they want to buy $\overline{c}$ without first trying $\underline{c}$. This allows the manufacturer to increase the wholesale price for $\underline{c}$ to the point where the period 1 participation constraint is binding ($\underline{w}^d_{t=1} = 2(1-h)v - s$) leading to a profit of $\pi_{3b} = 2v - s - \underline{c} - h(v + \overline{c})$.

A comparison between π_1, π_2 and π_{3b} yields

Proposition 5 *If policies 1,2 and 3b are available to the monopolistic manufacturer then profit maximization and vertical restraints will ensure full efficiency.*

Proof. Easily verified by comparing π_1, π_2 and π_{3b}. ■

Referring back to Figure 1, the manufacturer would follow Policy 1 in Region A, Policy 2 in Region B and Policy 3b in Region C.

Although Policy 3b is feasible in our simple model, it is a policy that only makes sense in a static context with a fixed population. Up to now, this simplifying assumption did not play any role for our results. But here it definitely does. In a more elaborate model, we envision the market as operating over time without beginning or end. In any period, those consumers who were successfully treated – or, whose problem is left untreated for two periods – depart from the market and there is a flow of new consumers into the market. In such an elaborate model, Policy 3b is obviously infeasible.

Is there another alternative to Policy 3a (or Policy 3b, respectively)? The following strategy is a candidate for a profit maximizing option

3c Sell only $\underline{c}$, and sell it through discounters only ($\underline{w}^e = \overline{w}^e = \overline{w}^d = \infty$); set the price $\underline{w}^d$ such that all consumers buy the minor treatment immediately.

If the problem is left untreated, a consumer incurs cost $2v$, if the consumer buys $\underline{c}$ from a discounter, he incurs cost $s + \underline{w}^d + 2hv$. Thus, the maximal feasible wholesale price for $\underline{c}$ is $\underline{w}^d = 2v(1-h) - s$ leading to a profit of $\pi_{3c} = 2v(1-h) - s - \underline{c}$.

The use of Policy 3c leads to a new kind of inefficiency, namely, that some customers do not receive sufficient treatment.

Proposition 6 *If Policy 3b is infeasible, then the manufacturer will employ Policy 1 iff* $h \in [\frac{c}{2v-(\overline{c}-\underline{c})}, 1 - \frac{c}{\overline{c}-\underline{c}}]$, *Policy 2 iff* $h > \max\{1 - \frac{c}{\overline{c}-\underline{c}}, \frac{\overline{c}-\underline{c}}{2v}\}$, *and Policy 3c iff* $h < \min\{\frac{c}{2v-(\overline{c}-\underline{c})}, \frac{\overline{c}-\underline{c}}{2v}\}$. *Thus, there exist (i) parameter constellations for which consumers inefficiently visit an expert instead of blindly buying* $\underline{c}$ *from a discounter; (ii) parameter constellations for which consumers inefficiently immediately receive the major treatment instead of first receiving the minor and if necessary the major treatment; and (iii) parameter constellations for which consumers are inefficiently left untreated if the minor treatment fails.*

Proof. Easily verified by comparing π_1, π_2 and π_{3c}. ■

Figure 3 illustrates the result. In area $A" \cap C$ consumers should blindly buy $\underline{c}$ from a discounter and if $\underline{c}$ fails they should then get $\overline{c}$. Now, they visit an expert. As compared to the first best policy this leads to an efficiency loss of $c - h(v + \underline{c})$. In area $B" \cap C$ customers are overtreated by receiving always a high quality treatment, even though the efficient policy is to sell first the minor treatment and – only if the minor treatment fails – the major treatment. As compared to the first best policy this leads to an efficiency loss of $\overline{c} - \underline{c} - h(v + \overline{c})$. In area $C"$ all consumers should blindly buy $\underline{c}$ from a

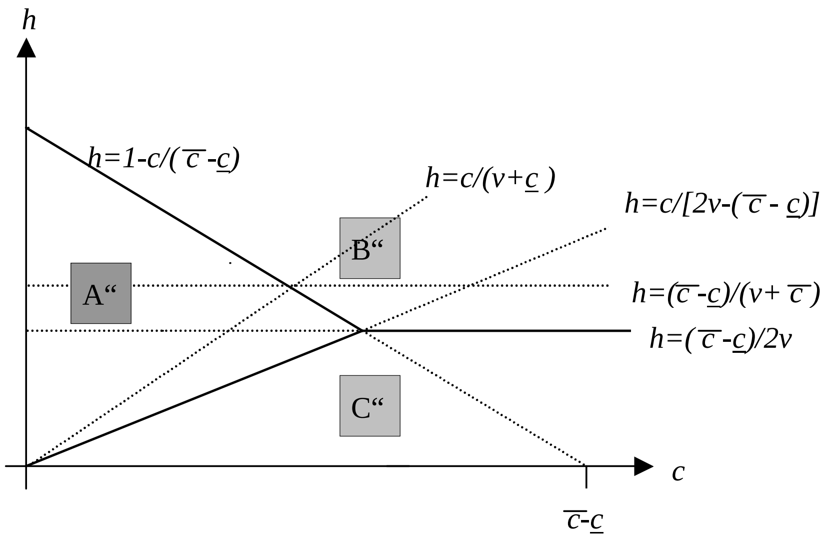

Figure 3: If Policy 3b is infeasible, the monopolisitc manufacturer employs Policy 1 in Region A”, Policy 2 in Region B”, and Policy 3c in Region C” (dotted lines are the – efficient – borders of regions A, B and C in Figure 1).

discounter and if $\underline{c}$ fails they should then get $\overline{c}$. Now, they have no possibility to buy $\overline{c}$. In comparison to the first best policy this leads to an efficiency loss of $h(v - \overline{c})$.

6.2 Heterogeneity in the Expected Cost of Efficient Treatment

In the previous subsection our focus was on the basic model with homogeneous consumers. In Section 5 we have seen that the inefficiencies of the homogeneous consumers case amplify if consumers are heterogeneous. A natural question therefore is whether vertical restraints can help to overcome these additional problems. For the case of heterogeneity in s the answer is straightforward. The manufacturer can simply rule out the existence of discounters and she has incentives to do so whenever consumers efficiently should visit an expert. Hence, the potential inefficiencies caused by heterogeneity in s will be avoided by the manufacturer. From the analysis of the previous subsection we also know that the magnitude of s does not affect the optimal policy of the manufacturer. Thus, the results of Subsection 6.1 also apply for the case of heterogeneity in s.

With respect to heterogeneity in h the manufacturer faces the problem that the wrong segment of the market visits an expert. A first question of interest is whether a benevolent manufacturer would have means to correct, or at least ameliorate, this adverse selection problem. The answer turns out to be yes; a benevolent manufacturer can indeed choose prices such that (i) she makes a non-negative profit; and (ii) the first best allocation prevails on the market. We record this result as Proposition 7. In the result reference is made to $\widehat{h}$. This variable stands for the average h of those consumers who efficiently should visit an expert. That is, if we define $h_1 = \frac{c}{v+\underline{c}}$ and $h_2 = 1 - \frac{c}{\overline{c}-\underline{c}}$ then $\widehat{h} = E(h\,|h_1 \leq h \leq h_2)$.

Proposition 7 *Wholesale prices satisfying* $\underline{w}^d = \underline{c}, \overline{w}^d = \overline{c}$, $\underline{w}^e = \underline{c} - \frac{c\widehat{h}}{1-\widehat{h}}$ *and* $\overline{w}^e = \overline{c} + c$ *ensure an efficient market outcome at no loss to the manufacturer. These wholesale prices lead to consumer prices* $\underline{q} = \underline{c}, \overline{q} = \overline{c}$, $\underline{p} = \underline{c} + c$ *and* $\overline{p} = \overline{c} + c$.

Proof. First notice that the consumer prices listed in the proposition reflect the real resource cost of each policy. Thus, consumers faced with those prices will behave efficiently.[17] Also notice that the quoted wholesale prices lead to the listed consumer prices. This follows from the analysis in Section 3. Finally notice that the quoted wholesale prices lead to zero profit for the manufacturer. For the prices charged from discounters this is obvious. With the wholesale prices charged from experts the manufacturer earns a profit of c on all $\overline{c}$ treatments sold, leading to an expected gain of $\widehat{h}c$; and she loses $\frac{c\widehat{h}}{1-\widehat{h}}$ on all $\underline{c}$ treatments sold, amounting to an expected loss of $(1-\widehat{h})\frac{c\widehat{h}}{1-\widehat{h}} = c\widehat{h}$. Thus, gains and losses cancel out in expectation. ■

[17] That free-riding on an expert's advice is unprofitable follows from $s \geq \overline{c} - \underline{c}$ and $0 < c < \frac{(\overline{c}-\underline{c})(v+\underline{c})}{(v+\overline{c})}$ (we made these assumptions in our analysis of heterogeneity in h) implying $s > c$.

Given that a benevolent manufacturer could, without losses, employ vertical restraints such that full efficiency prevails on the market, a next question of interest is, whether a profit-maximizing manufacturer will behave in that way. Not surprisingly, the answer turns out to be no. To show this, we first prove that a profit-maximizing manufacturer will never choose $\underline{w}^d$ and $\overline{w}^d$ in such a way that some consumers first try $\underline{c}$ and, if $\underline{c}$ fails, then buy $\overline{c}$.

Lemma 1 *A profit-maximizing manufacturer never charges wholesale prices such that some consumers first buy* $\underline{c}$ *from a discounter and, if* $\underline{c}$ *fails, then buy* $\overline{c}$ *from a discounter.*

Proof. Assume the opposite; that is, assume that there is a strictly positive measure of types that employs this strategy. First consider the situation where only discounters are active on the market. If some consumers employ the stated strategy then there must exist a critical type $\hat{h} \in (0, 1]$ such that the consumers in $[0, \hat{h}]$ are those who use this strategy, while the rest immediately buys $\overline{c}$ (here note that we allow for $\hat{h} = 1$). This follows from the fact that consumers' expected utility is strictly decreasing in h under the former strategy while it is type-independent under the latter. If consumers in $[0, \hat{h}]$ are expected to buy $\overline{c}$ if $\underline{c}$ fails, then the price for $\overline{c}$ can be at most v. At this price those consumers who immediately buy $\overline{c}$ have a rent of $v - s$. To give consumers who first try $\underline{c}$ at least the same rent the wholesale price $\underline{w}^d$ must satisfy $\underline{w}^d \leq (1 - 2\hat{h})v$. With $\overline{w}^d \leq v$ and $\underline{w}^d \leq (1 - 2\hat{h})v$ the manufacturer's profit on consumers in $[0, \hat{h}]$ is at most $(1 - \hat{h})v - \underline{c} - \hat{h}\ \overline{c}$ which is strictly less than v; and her profit on consumers in $(\hat{h}, 1]$ is at most $v - \overline{c}$ which is also strictly less than v. Now, an alternative policy for the manufacturer is to set $\underline{w}^d = \infty$ and $\overline{w}^d = 2v - s$. With this policy her profit per consumer is $2v - s - \overline{c}$ which is strictly more than v since $v - s > \overline{c}$. Hence, the original policy of the manufacturer must have been strictly dominated. The argument for the situation where there exists a segment of consumers who visits an expert is similar. The main point is that if the manufacturer charges wholesale prices such that some consumers buy $\overline{c}$ after having tried

$\underline{c}$ then all consumers must earn a rent of at least $v - s$. If $\underline{w}^d$ is set to ∞ and $\overline{w}^d$ is set to $2v - s$ then $\underline{w}^e$ and $\overline{w}^e$ can be adjusted in such a way that the manufacturer makes more profit on each single consumer. ■

Lemma 1 already tells us that profit-maximization by the manufacturer is in conflict with efficiency. We proceed by characterizing the market equilibrium implemented by the manufacturer. In our result reference is made to h_1 and to h_2. These variables are as defined in the paragraph preceding Proposition 7; that is, h_1 is the h such that in an efficient solution all consumers in $[0, h_1)$ should first try $\underline{c}$ and, if necessary, then buy $\overline{c}$; and h_2 is the h such that consumers in $[h_1, h_2]$ should visit an expert and consumers in $(h_2, 1]$ should immediately buy $\overline{c}$ from a discounter. In the result reference is also made to $\widetilde{h}_1$. This variable solves $\max \pi(\tilde{h}) = F(\tilde{h})[2v(1 - \tilde{h}) - s - \underline{c}] + \int_{\tilde{h}}^{h_2} [2v - s - c - \underline{c} - h(\overline{c} - \underline{c})] dF(h)$ and it satisfies $\widetilde{h}_1 < \frac{c}{2v - \overline{c} + \underline{c}} < \frac{c}{v + \underline{c}} = h_1$.[18]

Proposition 8 *A profit-maximizing manufacturer sets wholesale prices such that (i) consumers in* $[0, \widetilde{h}_1)$, *with* $\widetilde{h}_1 < h_1$, *buy* $\underline{c}$ *from a discounter and, if* $\underline{c}$ *fails, they leave the market; (ii) consumers in* $[\widetilde{h}_1, h_2]$ *visit an expert; and (iii) consumers in* $(h_2, 1]$ *buy* $\overline{c}$ *from a discounter.*

Proof. Assume that consumers who are indifferent between visiting an expert and visiting a discounter choose the efficient distribution channel (we will discuss this assumption in a footnote at the end of this proof) and consider the following wholesale prices $\underline{w}^d = 2v(1 - \widetilde{h}_1) - s$, $\overline{w}^d = 2v - s$, $\underline{w}^e = 2v - s - \frac{c}{1 - E(h | \widetilde{h}_1 < h < h_2)}$ and $\overline{w}^e = 2v - s$. Those wholesale prices imply consumer prices $\underline{q} = 2v(1 - \widetilde{h}_1) - s$, $\overline{q} = 2v - s$, $\underline{p} = 2v - s$ and $\overline{p} = 2v - s$. With these prices consumers' expected utility from buying $\underline{c}$ from a discounter is strictly decreasing in h while consumers' expected utility from visiting an expert and from buying $\overline{c}$ from a discounter is type-independent. Thus, since consumer prices are such that type $\widetilde{h}_1$ is exactly indifferent between buying $\underline{c}$ from a discounter and visiting an expert, consumers who face such prices

[18] The property $\widetilde{h}_1 < \frac{c}{2v - (\overline{c} - \underline{c})}$ can easily be verified by noting that the derivative of the objective function evaluated at $\frac{c}{2v - (\overline{c} - \underline{c})}$ is strictly negative.

will behave as stated in the proposition. Now we argue that the proposed wholesale prices are indeed optimal for the manufacturer. For consumers in $[h_1, h_2]$ and in $(h_2, 1]$ this is obvious They behave efficiently and get no rent. Thus, the manufacturer extracts the maximal feasible profit from them. Also, the expected price those consumers pay is type-independent and holds them exactly to their reservation utility. So, there is no negative side effect on consumers in $[0, h_1]$. Thus, $\overline{q}$, $\underline{p}$ and $\overline{p}$ are indeed optimal. Now consider consumers in $[0, h_1]$. It is obvious that in a profit-maximizing solution some of them will visit an expert while some others will buy $\underline{c}$ from a discounter. How does the manufacturer determine the cut-off point $\widetilde{h}_1$? Technically the answer is straightforward: She solves the maximization problem stated in the paragraph preceding the proposition. And economically? One possibility for her would be to choose the critical consumer where she makes exactly the same profit by serving this consumer through an expert and by selling him $\underline{c}$ through a discounter. This critical consumer is given by $h = \frac{c}{2v - \overline{c} + \underline{c}}$. Choosing this consumer's h as the cut-off value cannot be optimal, however. Why? Because at this point reducing $\widetilde{h}_1$ has only a second order effect on the profit made on consumers located at this point while it has a first order effect on the rent received by all consumers with a lower h. Thus, $\widetilde{h}_1 < \frac{c}{2v - (\overline{c} - \underline{c})}$ as claimed above.[19] ■

Let us summarize the findings of this subsection. First, we observed that the manufacturer could, in principle, choose prices such that full efficiency

[19]Notice that this proof relies on a tie breaking rule for consumers in $[\tilde{h}_1, 1]$. Why is such a tie breaking rule needed? Take $\overline{w}^d = 2v - s$ as given. To induce consumers in $[\tilde{h}_1, h_2]$ to visit an expert and consumers in $(h_2, 1]$ to buy $\overline{c}$ from a discounter consumer prices $\overline{p}$ and $\underline{p}$ must satisfy $\overline{p} \geq \underline{p}$ and $(1 - h_2)\underline{p} + h_2\overline{p} = 2v - s$. With $\overline{p} > \underline{p}$ consumers in $[\tilde{h}_1, 1]$ would have a strict incentive to do what they are supposed to do. However, setting $\overline{p} > \underline{p}$ implies a loss to the manufacturer because consumers with an h in $[\tilde{h}_1, h_2)$ would receive a rent compared to consumers in $[h_2, 1]$. So, equilibrium prices must satisfy $\overline{p} = \underline{p} = 2v - s$. But with $\overline{q} = \overline{p} = \underline{p} = 2v - s$ all consumers in $[\tilde{h}_1, 1]$ are indifferent between both distribution channels. Note, however, that this indifference problem can be solved at an arbitrarily small cost to the manufacturer by setting $\overline{p} = 2v - s + \epsilon$ and $\underline{p} = 2v - s - h_2\epsilon/(1 - h_2)$ with ϵ arbitrarily small.

prevails on the market. Next, we showed that a profit-maximizing manufacturer will not do so. She resorts to a policy such that a) consumers in $[0, \widetilde{h}_1]$ are inefficiently left untreated if the minor treatment fails, b) consumers in $(\widetilde{h}_1, h_1)$ inefficiently visit an expert instead of blindly buying $\underline{c}$ (and if $\underline{c}$ fails, then blindly buying $\overline{c}$) from a discounter, and c) consumers in $[h_1, 1]$ are efficiently served.

7 Conclusions

We have discussed a problem of double sided moral hazard that can be quite often observed in daily life. Whenever an expert can provide help to choose the appropriate quality of a good or service needed, there is scope, on the one hand, for the expert first to cheat on providing sincere (and costly) diagnosis and second to abuse her position and to sell to consumers the treatment that is most profitable for her; and, on the other hand, there is scope for consumers to cheat on experts by once having received her advice, buying the recommended good or treatment from some non-expert supplier.

We have shown that even if experts can charge for diagnosis, they will not do so in equilibrium, unless diagnosis can be observed and verified. Also, experts can not finance diagnosis costs by the mark-up for major treatments. If they would use either a diagnosis fee or a mark-up for major treatments to finance diagnosis costs, they would generate an incentive for themselves to refrain from diagnosis and to always provide major treatments. Such behavior would be expected by consumers and is therefore unprofitable.

The necessity to sell high quality without a mark-up and to provide diagnosis free of charge implies that diagnosis costs must be earned only through the mark-up on minor treatments. This has several implications. We have first studied a market where consumers are homogeneous and where the goods/treatments are provided by a competitive industry and sold by expert and discount (non-expert) sellers. In this setting there exist parameter constellations, where experts cannot survive competition by discounters even though selling through experts is efficient. This problem becomes even worse if consumer are heterogeneous in either switching costs or in the expected cost of efficient treatment. In the former case those consumers that switch easily increase the cost that has to be carried by consumers that are less inclined to switch. This cost-increase might set in motion a chain reaction like falling dominoes some additional consumers will free-ride, the mark-up will have to increase again, etc. As a consequence, a slight change in the composition of the population of consumers can completely unravel a market otherwise (efficiently) served by experts. If consumers are heterogeneous in expected cost of efficient treatment, the additional mark-up on minor compared to major treatments induces the wrong segment of the market to consult an expert. Given that diagnosis is free for consumers who need a high quality treatment, those consumers with a high propensity to need major treatments will (inefficiently) consult an expert and those consumers with a low propensity to need major treatments will (inefficiently) visit a discounter. This might increase the price experts must charge for minor treatments to such an extent that experts cannot survive competition by discounters.

We have also studied whether vertical restraints, such as retail price maintenance and minimum standards can overcome the inefficiencies involved. We have shown that wholesale prices could - in principle - be chosen in such a way that full efficiency prevails on the market. However, a profit-maximizing manufacturer will not do so. She resorts to a policy such that (i) some consumers inefficiently visit an expert instead of blindly buying a minor treatment from a discounter, and (ii) some consumers are inefficiently left untreated if the minor treatment fails.

References

[1] Bolton, P. and Bonano, G. (1988), "Vertical Restraints in a Model of Vertical Differentiation", *Quarterly Journal of Economics* **103**, 555-570.

[2] Bouckaert, J. and Degryse, H. (2000), "Price Competition between an Expert and a Non-Expert", *International Journal of Industrial Organization* **18**, 901-923.

[3] Dulleck, U. and Kerschbamer, R. (2005a), "On Doctors, Mechanics and Computer Specialists – Or, Where are the Problems with Credence Goods", Mimeo, Department of Economics, University of Vienna, forthcoming in the *Journal of Economic Literature.*

[4] Dulleck, U. and Kerschbamer, R. (2005b), "Experts vs. Discounters Competition and Market Unraveling When Consumers Do Not Know What They Need, Discussion Paper No. 3565, CEPR, London.

[5] Emons, W. (1997),"Credence Goods and Fraudulent Experts", *Rand Journal of Economics **28**,* 107-119.

[6] Emons, W. (2000), "Product Differentiation and Price Competition between a Safe and a Risky Seller", *Journal of Institutional and Theoretical Economics* **156**, 431-444.

[7] Emons, W. (2001), "Credence Goods Monopolists", *International Journal of Industrial Organization* **19**, 375-389.

[8] Glazer, J. and McGuire, T. (1996), "Price Contracts and Referrals in Markets for Services", Working Paper No 10/96, Institute of Business Research, Tel Aviv University.

[9] Klein, B. and Murphy, K. M. (1988), " Vertical Restraints As Contract Enforcement Mechanisms", *Journal of Law and Economics* **31**, 265-297.

[10] Marvel, H. and McCafferty, S. (1984), "Resale Price Maintenance and Quality Certification", *RAND Journal of Economics* **15**, 346-359.

[11] Pesendorfer, W. and Wolinsky, A. (2003), "Second Opinions and Price Competition Inefficiency in the Market for Expert Advice", *Review of Economic Studies* **70**, 417-437.

[12] Perry, D. and Besanko, M. K. (1991), "Resale Price Maintenance and Manufacturer Competition for Exclusive Dealerships", *Journal of Industrial Economics* **39**, 514-544.

[13] Pitchik, C. and Schotter, A. (1987), "Honesty in a Model of Strategic Information Transmission", *American Economic Review* **77**, 1032-1036; Errata (1988), *AER* **78**, 1164.

[14] Shaw, D. J. (1994), "Resale Price Maintenance as a Private Contract Enforcement Mechanism The Special Services Case", *Southern Economic Journal* **61**, 52-68.

[15] Sülzle, K. and Wambach, A. (2003), "Insurance in a Market for Credence Goods", forthcoming in *Geneva Papers on Risk and Insurance Theory.*

[16] Telser, G. (1960), "Why Should Manufacturers Want Fair Trade?", *Journal of Law and Economics* **3**, 86-105.

[17] Taylor, C. (1995), "The Economics of Breakdowns, Checkups, and Cures", *Journal of Political Economy* **103**, 53-74.

[18] Wolinsky, A. (1993), "Competition in a Market for Informed Experts' Services", *Rand Journal of Economics* **24**, 380-398.

[19] Wolinsky, A. (1995), "Competition in Markets for Credence Goods", *Journal of Institutional and Theoretical Economics* **151**, 117-131.

8 Appendix

In this appendix we show that Equation (7) yields at least one solution for $\widetilde{h}$.

Equation (7) is equivalent to

$$\frac{1-\widetilde{h}}{\widetilde{h}[1-E(h \,\big|\, h \geq \widetilde{h})]} = \frac{1}{x}, \tag{8}$$

where x is defined as $x = \frac{c}{(v+\underline{c})}$. Since $c < \frac{(\overline{c}-\underline{c})(v+\underline{c})}{(v+\overline{c})}$ and $v > \overline{c}$ we have $x < \frac{(\overline{c}-\underline{c})}{(v+\overline{c})} < \frac{1}{2}$. Thus, the right hand side of Equation (8) is strictly larger than two. To prove existence of a solution we now show (a) that the left hand side of Equation (8) converges to $+\infty$ if $\widetilde{h}$ converges to zero and (b) that it converges to two if $\widetilde{h}$ converges to one. Part (a) is trivial given that $E(h\,|h \geq 0) = E(h) \in (0,1)$. To show part (b) first notice

$$\frac{(1-\widetilde{h})}{[1-E(h\big|h\geq\widetilde{h}\big)]} = \frac{(1-\widetilde{h})}{1-\frac{1}{1-G(\widetilde{h})}\int_{\widetilde{h}}^{1} hg(h)dh} = \frac{(1-\widetilde{h})(1-G(\widetilde{h}))}{1-G(\widetilde{h})-\int_{\widetilde{h}}^{1} hg(h)dh} = \frac{(1-\widetilde{h})(1-G(\widetilde{h}))}{\int_{\widetilde{h}}^{1} g(h)dh-\int_{\widetilde{h}}^{1} hg(h)dh} =$$

$\frac{(1-\widetilde{h})(1-G(\widetilde{h}))}{\int_{\widetilde{h}}^{1}(1-h)g(h)dh}$. Thus, by applying l'Hôpital's rule twice we get $\lim\limits_{\widetilde{h}\to 1}\frac{(1-\widetilde{h})[1-F(\widetilde{h})]}{\int_{\widetilde{h}}^{1}(1-h)f(h)dh} = \lim\limits_{\widetilde{h}\to 1}\frac{-[1-F(\widetilde{h})]+(1-\widetilde{h})[-f(\widetilde{h})]}{-(1-\widetilde{h})f(\widetilde{h})} = \lim\limits_{\widetilde{h}\to 1}\frac{(1-F(\widetilde{h}))}{(1-\widetilde{h})f(\widetilde{h})}+1 = \lim\limits_{\widetilde{h}\to 1}\frac{-f(\widetilde{h})}{-f(\widetilde{h})+(1-\widetilde{h})f'(\widetilde{h})}+1 = 2$ (provided f is differentiable!). ■