THE JOURNAL OF INDUSTRIAL ECONOMICS 0022-1821
Volume LX September 2012 No. 3

AGENCY PROBLEMS AND REPUTATION IN EXPERT SERVICES: EVIDENCE FROM AUTO REPAIR*

HENRY S. SCHNEIDER†

Using a field experiment involving undercover visits to auto repair garages with a test vehicle, I first examine how asymmetric information between mechanics and motorists over auto repair service quality affects outcomes. I then examine whether reputation mitigates these problems via a matched-pair treatment in which undercover researchers appeared as either one-time or repeat-business customers. The results indicate that under and overtreatment are widespread, and that reputation via a repeat business mechanism does not improve outcomes significantly.

I. INTRODUCTION

IN MANY SERVICE MARKETS, SUCH AS AUTOMOBILE AND BOAT REPAIR, plumbing and roofing work, and many medical specialties, the seller of the service is also the expert who diagnoses how much service is needed. This dual relationship raises concerns about agency problems and inefficient levels of service provision in expert service markets (Darby and Karni [1973]).

A more recent theoretical literature, however, raises the possibility that expert service markets may operate efficiently. Wolinsky [1993], Emons [1997, 2001], Alger and Salanie [2006], and Dulleck and Kerschbamer [2006] identify equilibria in which expert service markets operate efficiently in a one-period setting.[1] Further, even when inefficient outcomes would occur in a one-period setting, an expert's concern for her reputation might limit inefficiencies in repeat-business settings. Klein and Leffler [1981] and Shapiro [1983] are seminal papers in the pure moral hazard reputation literature, while Kreps and Wilson [1982], Milgrom and Roberts [1982], and Kreps, Milgrom, Roberts and Wilson [1982] allow for similar out-

*I thank Steve Berry, Justin Fox, Don Green, Justin Johnson, Dean Karlan, Josh Lustig, Steve Nafziger, Philipp Schmidt-Dengler, Fiona Scott Morton, Harsha Thirumurthy, Michael Waldman, various seminar participants, and the Editor and anonymous referees for comments, and the Yale Institution for Social and Policy Studies for financial support. I also thank Michael Coughlin at Premier Subaru, Romana Primus at Whaling City Ford and Paul Trembley at East Rock Auto Repair for background information about the auto repair industry, and George Iny and his colleagues at the Automobile Protection Association for invaluable assistance in designing the experiment and interpreting the outcomes.

†Authors' affiliations: Johnson Graduate School of Management, Cornell University, Ithaca, New York 14853 U.S.A.
e-mail: henry.schneider@cornell.edu

[1] Many of these and other papers identify equilibria with inefficient outcomes as well.

comes in a framework that also includes adverse selection over sellers' types. More recent predictions of effective reputation mechanisms in expert services are given in Wolinsky [1993] and In-Uck Park [2005].

While I describe the literature on agency problems and reputation as it relates to expert service markets in more detail in Section II, it is apparent that the predictions of the theoretical literature are sensitive to the conditions assumed by the models (this sensitivity is discussed in Dulleck and Kerschbamer [2006] with respect to agency problems). The expert services literature predicts a wide range of outcomes regarding whether asymmetric information generates undertreatment, overtreatment, or simply overcharging, while the reputation literature predicts that reputation may mitigate, exacerbate or have no effect on the extent of problems.

The objective of the current study is to provide empirical results regarding several of these predictions in a commonly cited expert service market with information problems, the market for auto repairs. I first test for agency problems in the market. Upon finding that problems are pervasive, I then examine a repeat-business setting to test whether reputational factors have an effect on these problems.

I use data derived from 40 undercover garage visits I conducted during a field experiment (the 'experimental' visits) with the help of mechanics working with a Canadian public-interest group, the Automobile Protection Association (APA), and repair receipts from 51 undercover garage visits conducted by the Canadian group itself (the 'Canadian' visits).[2] During the garage visits, an undercover researcher posed as an ordinary motorist and submitted a test vehicle with a prearranged set of defects to garages for repairs. The mechanic was asked to thoroughly inspect the vehicle, diagnose its condition, make a repair recommendation, and provide a price estimate. To characterize the nature of agency problems, I then compare the observed outcomes with the repairs that I knew ahead of time were actually necessary.

The tests for agency problems involved defects that were (1) sufficiently simple to diagnose that they were designed to identify intentional overtreatment and neglect instead of competency; and (2) had treatments that were not subject to discretion regarding their type or necessity. That is, a mechanic who were acting as an agent for the motorists to maximize the motorist's indirect utility would not face ambiguity over how to treat the defects. These defects were a battery cable that was visibly loose at the battery terminal (located conspicuously at the front of the engine compartment) that caused an intermittent starting problem, which was the ostensible reason for visiting the garage and for which symptoms were explicitly

[2] See List [2006a] for a discussion of field experiments in economics.

described to the mechanic (the APA mechanics called this test 'fool-proof'); a low level of coolant that was again visible in the engine compartment, a check of which is a standard component of any basic vehicle inspection (for example, a safety or pre-purchase inspection) and which risked the life of the engine; and a missing taillight, a check of which is again standard in any basic vehicle inspection. Appendix A contains a photograph of the test vehicle and the locations of the defects.

The motivation for the intermittent starting problem is to measure the extent to which mechanics acted opportunistically to conduct and/or charge for unnecessarily costly repairs, of which there were several obvious possibilities including starter motor or battery replacement, instead of simply tightening the loose cable. The motivation for the low coolant and missing taillight is to identify whether mechanics would exert effort to identify the defects, which are trivial to correct once discovered, but require modest but still costly effort.

To assess the ability of reputational incentives to mitigate agency problems, I randomly assigned mechanics to receive one of two procedures. In one, I appeared as a one-time customer, stating that I was moving away and having moving boxes visible in the back of the car. In the other, I appeared as a possible repeat customer by providing a home address near to the garage, and suggesting that I was seeking a local mechanic for an ongoing relationship. I then estimate the effect of this repeat-business procedure on the number of legitimate defects discovered, the diagnosis fee, the repair recommendation and the repair price. I repeat this test using the Canadian data, relying on variation in whether the undercover researcher represented possible repeat business that was unknowingly generated by the Canadian group during their visits.

I now summarize the outcomes. First, overtreatment was common. I find that mechanics recommended completely unnecessary repairs in 33 per cent of the experimental visits and 27 per cent of the Canadian visits. This overtreatment appeared as repairs that were significantly more costly than necessary to correct the defects, or that were unrelated to the defects altogether. For example, a common overtreatment for the intermittent starting problem was to replace a healthy starter motor, which is a costly repair, instead of to tighten the loose battery cable, which was the necessary but also low-cost repair. In contrast, overcharging, defined as charging for parts or labor that were not actually provided, appeared to be relatively limited. Second, undertreatment was very common. Seventy-five per cent of the field experiment visits were characterized by serious undertreatment, which I define as the mechanic's missing the loose battery cable or the low level of coolant, the two simplest and most important defects to diagnose. Additionally, the missing taillight was undetected in 87 per cent of visits. This first set of results demonstrates that service quality on average is quite low and appears to be far from an efficient level, and is consistent with

outcomes under an important moral hazard faced by mechanics over their effort and repair choices.[3]

I next test whether reputation in the form of motorists' representing repeat versus one-time business generates higher quality service and mitigates the moral hazard problem. Here I find no evidence that motorists representing repeat business receive different repair recommendations, repair prices or diagnosis quality versus one-time motorists, and hence I find no evidence that reputation limits inefficiencies: during repeat-business visits, the average quality of diagnoses continued to be poor, and the type and amount of repairs remained inconsistent. A repeat-business effect was evident, however, for the diagnosis fee. This fee was $37.70 on average when I appeared as repeat business, while it was $59.75 when I appeared as one-time business. The difference between these amounts represents three-quarters of a standard deviation of the diagnosis fee ($p = 0.02$). Results from the Canadian visits are very similar.

The diagnosis fee result validates the one-time versus repeat-business procedures in the sense that it shows that mechanics observed and responded to this variation. This result also suggests that mechanics do respond to repeat business incentives when motorists can easily evaluate the favorability of service. In the case of the diagnosis fee, motorists learn this fee after the inspection, and at this point can evaluate its favorability in deciding whether to return for repeat business.[4] In contrast, it is very hard for motorists to evaluate the quality and necessity of most types of repairs, even after service has been provided, and here we observe no reputation effects. This contrast is consistent with a complementarity between customer learning and reputation effects.

The importance of customer learning is also suggested by the difference in correction rates between the one problem where the quality of service may be less difficult for motorist to evaluate after the visit, which is the intermittent starting problem and whether it was corrected, and the remaining defects, which are harder for motorists to evaluate. The intermittent starting problem is corrected in 78 per cent of visits (once unnecessary repairs that would have incidentally corrected the loose battery cable are included). In contrast, the low coolant level and missing taillight, which are harder to attribute to low mechanic effort (e.g., the driver may not know if

[3] While the defects were designed to be among the simplest a mechanic would face, it is possible that some low quality service is due to basic incompetence. The interpretation of the inefficiency would be somewhat different under incompetence: Moral hazard may be driving the outcomes, but would occur via garage owners' decisions regarding whether to hire competent mechanics, and perhaps mechanics' decisions as to how much training to obtain. I discuss this possibility and provide additional evidence that mechanic skill is unlikely to explain the results in Section VI.

[4] For example, motorists can evaluate that a fee of $0 is favorable, while a fee over $100 is unfavorable.

the defects arose only *after* the visit and hence may not be able to attribute their presence to low mechanic effort), are corrected in 28 and 13 per cent of visits.[5]

Finally, I note that, while service quality was often quite low, there are apparently enough mechanics providing sufficiently good quality to prevent market failure in this segment of the industry. For example, the intermittent starting problem *was* corrected in most visits.

The current paper provides a detailed view of moral hazard and reputation in a real-world expert service market, and the results point to information problems as an important determinant of market outcomes in this context. One qualification is that the focus of the reputation test is specifically on a repeat-business mechanism. The test does not address whether reputational factors affect mechanic behavior outside of a one-time versus repeat-business comparison. For example, some mechanics might always act in the interest of motorists while others might always act opportunistically regardless of motorist appearance. A second qualification is that if most mechanics lack the skill to conduct basic vehicle inspections, then even well intentioned mechanics could not provide better service to repeat customers, and again no reputation effect would be apparent. I discuss why this latter possibility is unlikely in Section VI.

The paper proceeds as follows. Section II describes the literature. Section III describes the Canadian and field experiment data. Sections IV and V describe the results for agency problems and reputation. Section VI discusses the results and alternative explanations. Section VII concludes.

II. RELATION TO LITERATURE

Given the dual role of an expert as diagnoser and seller of service, a primary question of the theoretical literature is whether expert service markets operate efficiently. The seminal paper of Darby and Karni [1973] raises the possibility of inefficient overtreatment. This dual role, however, does not necessarily imply overtreatment. If the buyer cannot observe which treatment was provided, and overtreating is costly, experts may instead charge for the high-cost treatment but provide the low-cost treatment. Pitchik and Schotter [1987], Wolinsky [1993], and Taylor [1995] predict variations of this outcome. Emons [1997, 2001], Dulleck and Kerschbamer [2006], and Alger and Salanie [2006] assume buyers observe treatment type after

[5] There is no difference in the starting-problem correction rate between one-time versus repeat-business treatments even though we might expect to see an effect here given its easier observability for motorists. However, this may simply reflect the already-high correction rate in the one-time treatments (78 per cent). This high correction rate in the one-time treatments could reflect that failure to fix this problem may be immediately apparent to the motorist, and hence might prevent the mechanic from collecting fees for the current service.

service, such that experts must conduct treatment in order to induce excess demand, and these papers identify equilibria with inefficient overtreatment.[6]

Emons [1997, 2001], Alger and Salanie [2006], and Dulleck and Kerschbamer [2006], also identify equilibria where expert service markets operate efficiently: Experts are honest in order to maximize consumer surplus, in the competitive case to survive, while in the monopoly case to maximize profits.[7] Dulleck and Kerschbamer [2006] generalize the credence goods problem to encompass many of these models by varying several key assumptions, identifying the conditions required for an efficient outcome.[8]

The papers above assume that experts diagnose problems perfectly thereby excluding a role for expert skill and effort. Pesendorfer and Wolinsky [2003] allow experts to choose whether to exert a costly hidden effort that is required to accurately diagnose the problem. In their model, customers incentivize experts to provide high quality diagnoses by obtaining second opinions. However, if few experts provide accurate diagnoses, the disciplining mechanism fails.

At least a dozen empirical studies test for overtreatment for physicians' services. These studies provide mixed but generally supportive evidence for agency problems. McGuire [2000] and Gaynor and Vogt [2000] review much of this literature, highlighting challenges to identification. Studies in other expert service markets generally find evidence of agency problems, including Ellison and Chevalier [1999] on mutual fund managers, Iizuka [2007] on physicians in the prescription drug market, Levitt and Syverson [2008] on real estate agents, and Schneider [2010] and Balafoutas, Beck, Kerschbamer and Sutter [2011] on taxi drivers. Dulleck, Kerschbamer and Sutter [2011] find significant agency problems in a laboratory experiment with college students in a credence-goods setting under some conditions. Beck, Kerschbamer, Qiu and Sutter [2009] compare the behavior of student subjects versus actual mechanics (Austrian car mechanic apprentices in training) that are brought into the laboratory setting, and find that the latter exhibit more amplified effects.

[6] Fong [2005] further identifies an important role for buyer heterogeneity in generating overtreatment, while Emons [1997] highlights the importance of capacity, such as that excess capacity may generate overtreatment while insufficient capacity may generate undertreatment.

[7] Wolinsky [1993] finds equilibria with additional mechanisms for generating honest diagnoses. In one, experts specialize such that some experts only treat minor problems while others only treat major problems, thereby eliminating incentive to overtreat minor problems. In another, all treatments have the same price, which also eliminates incentives to overtreat. Neither mechanism appears to occur in the segments of the market examined here.

[8] Dulleck, Kerschbamer and Sutter [2011] provide results from a laboratory experiment with college students to examine the role of these assumptions. Reputation is also investigated with some conditional effects being found.

Klein and Leffler [1981] and Shapiro [1983] are seminal papers on reputation: Buyers observe quality only after purchase, and return for repeat business if quality is sufficiently high. However, these models sustain a wide range of equilibria including bad outcomes. Mixed models that also allow for adverse selection over seller type, such that some sellers are strategic while others always act either in the interest or a gainst the interest of buyers, provide more concrete predictions. Kreps and Wilson [1982], Milgrom and Roberts [1982] and Kreps, Milgrom, Roberts and Wilson [1982] assume the existence of good-type sellers who always act in the interest of buyers, and strategic sellers who imitate good sellers in attempts to win a good reputation, thereby limiting inefficiency. Mailath and Samuelson [2001] assume instead the presence of bad-type sellers who always recommend overtreatment, and strategic sellers who wish to separate from bad sellers. Mailath and Samuelson find now that reputation may fail to generate efficient treatment.[9] Note that these predictions assume buyers observe quality after the transaction.

Models of reputation where service quality is not observed after the transaction also give varied predictions. Wolinsky [1993] and In-Uck Park [2005] find that an expert's concern for her reputation may limit inefficiencies. However, the 'bad reputation' results of Ely and Valimaki [2003] and Ely, Fudenberg and Levine [2008] predict that some experts may undertreat in order to separate from bad mechanics who always recommend expensive treatment, thereby generating inefficiently low levels of treatment.

Empirical work on experience and credence goods generally finds evidence of reputation effects, including in Hubbard [1998, 2002] on vehicle inspections, Banerjee and Duflo [2000] on the customized software industry in India, and Jin and Leslie [2009] on restaurant hygiene practices. List [2006b] finds that reputation induces sports-card dealers to provide higher quality products, but only when an explicit quality-grading system is available, thereby highlighting a complementarity between reputation and a buyer's ability to evaluate quality after purchase. Laboratory evidence on reputation is mixed. For example, Grosskopf and Sarin [2010] find that reputation tends to improve efficiency (in contrast to the bad reputation predictions mentioned above), though to a lesser degree than predicted by some theory, and Bolton, Katok and Ockenfels [2004] also find benefits of a reputation mechanism.[10] However, Dulleck, Kerschbamer and Sutter [2011] find that reputation has limited ability to address agency problems in a credence goods setting.

[9] Bar-Isaac and Tadelis [2008] describe the reputation literature in more detail.

[10] See Grosskopf and Sarin [2010], p. 2189, for references on experimental studies of related games, and Palfrey and Prisbey [1996], Andreoni and Croson [2008] and others on identifying reputation from altruism and other effects.

III. THE DATA

I analyze two sets of data based on field visits to garages by undercover researchers. I describe the data from the field experiment first and the Canadian visits second.

III(i). *Description of the Field Experiment*

Mechanics and researchers associated with the Canadian public-interest group, APA, provided guidance in preparing the test vehicle and implementing the experiment. The test vehicle was rigged with a loose battery cable designed to cause intermittent starting failure. This was the ostensible reason for visiting the garages. During each visit, I also told the mechanic that the vehicle was purchased recently and requested a thorough inspection to uncover any additional problems. I scheduled an appointment by phone, at which time I was usually asked for a description of the symptoms of the defect, the make and model of the vehicle, a name and phone number and, on three occasions, a home address. The script for this call is in Appendix B.

I generate true exogenous variation in the importance of reputation by randomly assigning mechanics to treatments in which reputation is either less or more important. I also collected garage and visit characteristics, such as garage size, mechanic certifications, arrival time, weather, and researcher appearance, to limit the risk of systematic differences in unmeasured customer and garage characteristics between one-time and repeat-business visits.

Upon visiting the garage, during the low-reputation procedure, I said I was moving to Chicago (from Connecticut) in two weeks and wanted the car examined for problems before the trip.[11] During the high-reputation procedure, I said I was moving in nearby and wanted the car examined for problems before traveling round-trip to Montreal in two weeks. Chicago and Montreal were chosen as destinations because the round-trip distance from Connecticut to Montreal is approximately equal to the one-way distance from Connecticut to Chicago. The exact scripts are provided in Appendix B. The low-reputation appearance was reinforced by placing two bags of U-Haul foam moving peanuts, 10 flattened U-Haul boxes, a push cart, an air conditioner box, a DVD-player box, a Dell computer box, and a microwave box in the car. During high-reputation visits, the inside of the car was bare.

[11] Upon arrival to the garage, the mechanic usually asked for the car's symptoms and service history, a telephone number, and sometimes a home address. In all but one visit, the inspection fee was only specified after the diagnosis was made. I did not request the inspection fee prior to leaving the car for inspection. However, note that such estimates are not binding over the approximately $100 range over which inspection prices vary, as mechanics can increase this price after the inspection is conducted by claiming small repairs or extra tests.

I paid the garages for the diagnostic inspection. If repairs were recommended, I told the mechanic I would consider having them conducted and call the next day if I wanted them done. After each visit, I visually inspected the engine compartment of the vehicle, and when necessary, applied a very light layer of grease and dirt to the battery to ensure a consistent appearance.

III(ii). *Test Vehicle, Undercover Researcher and Garages*

The test vehicle was a 1992 Subaru Legacy L Wagon. The vehicle had 141,000 miles and was thirteen years old at the beginning of the experiment, accumulated 4,000 additional miles during the experimental period, and had an appearance one would expect of a well-maintained thirteen-year-old car.[12] Photographs of the vehicle and engine compartment are in Appendix A.

Prior to data collection, the vehicle received thorough inspections from two APA mechanics who documented the condition of all of the car's parts and made judgments about whether the defects required immediate repair or just monitoring. In addition to the loose battery cable, four defects required immediate attention: a low level of coolant, a missing taillight, misfit and worn spark plug wires, and a one-inch exhaust pipe leak.[13] Other parts that displayed some wear but were still in good working condition, which the APA mechanics indicated required monitoring but not immediate attention, were a slightly weak alternator that still effectively charged the battery, an exhaust system with a moderate amount of rust along the center pipe and muffler, an unknown condition of the timing belt, moderately-worn shock absorbers, and two moderate oil leaks from the engine. The car's remaining parts were judged to be in very good condition. The undercover researcher (the author) was a 31-year-old Caucasian male at the time, and conducted all visits wearing khaki trousers and a polo shirt.

The study involves independent auto repair shops in four towns in two Connecticut counties. I visited only independent shops and not chains and franchises since mechanics at independent shops, who are typically the

[12] New license plates with a tag number corresponding to a June 2005 registration were installed prior to data collection to add credibility to the script that I had just purchased the car and moved to the area. A car may have new license plates if the researcher were moving away (consistent with the one-time business treatment) for many reasons. For example, if the plates were replaced because of damage.

[13] To maintain the appearance of low coolant throughout the experiment, I emptied the coolant overflow tank before each garage visit. Insufficient coolant, especially during the summer when the experiment was conducted, can cause the engine to overheat, and risks the life of the vehicle. One spark plug wire did not fit properly into the engine block, which allowed debris and rain water to enter on top of cylinder head, and could cause engine misfiring and corrosion. The exhaust pipe leak was located near the front of the center pipe beneath the driver's seat, which could allow fumes to enter an open window if the car was idling.

TABLE I
CHARACTERISTICS OF GARAGES IN EXPERIMENTAL VISITS

	One-time	Repeat	Total
Total	20	20	40
Gas station on premises	6	4	10
ASE certification	9	8	17
NAPA/AC Delco association	5	8	13
Garage size (number of active bays)	2.4	2.8	2.6
Heat index > 90F	9	4	13
Days to first appointment	1.7	2.0	1.8
Arrival time in hours	10:35 AM	11:10 AM	10:52 AM
City 1	7	7	14
City 2	4	6	10
City 3	3	4	7
City 4	6	3	9

Note: The table provides characteristics of the garages visited during the field experiment. One-time indicates the undercover researcher appeared as one-time business. Repeat indicates he appeared as possible repeat business.

owners themselves, interact directly with the customers, which ensures the mechanic is aware of the repeat versus one-time business status. In contrast, chain and franchised garages sometimes employ a counter employee who acts as a go-between for the customer and mechanic.[14]

Garage selection involved choosing towns in Connecticut, selecting garages that appeared in a Google maps search of that town, choosing a home address within 0.7 miles of the garage to provide to the mechanic during the visit, and matching garages in pairs based on similar characteristics. To make the subject sample as homogeneous as possible in order to improve test power, AAA-approved garages, auto body and oil change shops, and garages that sold cars on the lot were excluded.[15] Appointments were scheduled, and then garages were randomly assigned to receive the high or low-reputation treatment. While some heterogeneity remained, Table I shows that garage and visit characteristics were reasonably well balanced between groups.

The forty garages represent about one-third of all businesses registered with the Connecticut Department of Motor Vehicles as General Auto Repairers in these towns, though the garages represent a higher fraction of the independent garages in these towns since the General Repairers cat-

[14] Independent shops also depend more directly on individual relationships with customers for repeat business and referrals, as opposed to chains and franchises, which face a more complicated set of incentives from factors such as common branding and advertising across stores.

[15] Excluding AAA-approved shops was a natural choice since less than 10 per cent of independent garages in the two counties were AAA-certified for auto repair (approximately 25 garages). Pre-2003 APA visits provided no evidence that AAA-certified garages provide better service: Twelve garages visited by the APA were certified by the CAA (the Canadian equivalent of AAA), and overcharging and overtreating were slightly more frequent at these shops.

egory also includes dealerships and chain garages.[16] The four towns have similar sizes and socioeconomic characteristics. Their populations are 40,000–60,000 residents each, and are 65–80 per cent white and 10–20 per cent African-American. Median household incomes in these towns are between $40,000 and $55,000. Connecticut has significant socioeconomic diversity, with some of the wealthiest towns in the U.S. in the southwestern region near New York City (e.g., median household income in Darien, Connecticut, is over $150,000), and some of the poorest areas of the country in several of the major metropolitans areas (e.g., median household income in Hartford, Connecticut, is approximately $26,000). The garages are not in either of these regions, but instead in towns with socioeconomic characteristics that are more representative of the state as a whole.

III(iii). *Description of the Canadian Data*

I also constructed a data set from service receipts provided to me by the APA from a hidden-camera investigation of Canadian auto repair shops that was aired on the Canadian television network, CTV. The APA conducted 51 undercover visits to garages in Montreal, Toronto, Calgary and Vancouver during 2003. Of these garages, 23 were company-owned chains, 23 were franchises, four were independent shops, and one belonged to a car dealership. During each visit, the undercover researchers presented a vehicle with a loose battery cable, a defect that causes intermittent starting failure, and was plainly visible in the front of the engine compartment, easy to diagnose and fix with equipment that is standard at all garages, and because of its simplicity, designed to test for overtreatment and overcharging and not competency. At the start of each visit, the researchers told the mechanic the vehicle was purchased recently, and requested a general inspection to diagnose any additional problems and recommend necessary repairs. The test vehicle was an off-warranty five-year-old Dodge Caravan with approximately 50,000 miles. All of its serviceable parts were new or in excellent condition, and prior to each visit, APA mechanics inspected the vehicle to ensure their quality.[17]

The undercover researchers were a male-female couple except for four visits in Vancouver, when the couple was two females.[18] All of the researchers were Caucasian in their late thirties to early fifties. They consented to any repairs that were recommended by the mechanics, and for parts that could reasonably be returned, requested back the parts that the mechanic

[16] There are approximately 1,700 registered General Repairers in and around 169 incorporated towns in Connecticut.

[17] Any parts that were not in excellent condition were replaced prior to their visits, including the battery, fuel filter, spark plugs and wires, brake rotors, and tires, and the vehicle was transported between cities by railroad.

[18] The outcomes of the four female-couple visits were unexceptional.

TABLE II
CHARACTERISTICS OF GARAGES IN CANADIAN VISITS

	One-time	Repeat	Total
Total	20	29	49
Chain	10	13	23
Franchise	8	13	21
Independent	1	3	4
Dealer	1	0	1
Montreal	0	13	13
Toronto	0	16	16
Calgary	8	0	8
Vancouver	12	0	12

Note: The table reports characteristics of the Canadian visits. One-time indicates the undercover researchers appeared as one-time business. Repeat indicates they appeared as possible repeat business. Two visits are excluded from the table because the province of the license plate was inconsistent with the home address provided by the researcher to the garage, such that the repeat versus one-time appearance was ambiguous.

removed from the car, which APA mechanics re-installed into the vehicle before the next visit.[19] For each service receipt, I itemized the inspection price, repair price, various garage characteristics, and any recorded information provided to the garage by the undercover researcher such as the home address that was provided. I adjusted all prices to 2005 U.S. dollars.

In 29 APA visits to garages in Montreal and Toronto, APA researchers provided an address that was in the same city and had license plates that corresponded to the Canadian province of the visit. During 20 visits to garages in Calgary and Vancouver, the researchers gave an out-of-province address, had out-of-province license plates, and stated that they were traveling through the area on vacation.[20] These one-time and possible repeat business patterns were not deliberate choices of the researchers (and, in fact, went unnoticed as remarkable until our conversations), but merely reflected their true cities of residence. Table II describes the garages visited by APA researchers, and shows that the garage characteristics are balanced between one-time and repeat business visits.[21]

IV. EVIDENCE ABOUT AGENCY PROBLEMS

In this section, I report the patterns of service quality provided by mechanics. I see the contribution of these results primarily as revealing the low

[19] The researchers requested that the mechanic return the part after recommendations were made but before they were completed. Thus, this request should not affect recommendations.

[20] The province of the license plate is inconsistent with the provided home address for two of the visits to franchise garages in Calgary. These observations are excluded in the repeat-business analysis.

[21] One may worry that this one-time versus repeat-business measure might be confounded with city effects. That is, perhaps mechanics in Montreal and Toronto simply provide lower quality service on average than mechanics in Calgary and Vancouver. First, I have no reason to believe this is the case. Second, this question provides additional motivation for the random treatment assignments in the field experiment.

TABLE III
LEGITIMATE DEFECTS DISCOVERED DURING EXPERIMENTAL VISITS

	One-time		Repeat		Total	
	Discover	Repair	Discover	Repair	Discover	Repair
Loose battery cable	14	14	13	13	27	27
Low coolant	6	6	5	5	11	11
Missing taillight	3	3	2	2	5	5
Misfit plug wires	4	1	5	4	9	5
Exhaust pipe leak	2	1	4	2	6	3
Rusted muffler	8	4	8	2	16	6
Weak alternator	2	1	5	2	7	3
Timing belt service	10	4	11	5	21	9
Worn shocks	0	0	1	0	1	0
Oil leaks	0	0	2	0	2	0

Note: The table shows the number of visits in which each legitimate defect was discovered (Discovered) and recommended to be repaired (Repair) during the twenty one-time and twenty repeat-business experimental visits. Repair is sometimes less than Discover since mechanics may recommend a wait-and-see approach. The top and bottom panels indicate parts requiring and not requiring immediate attention, respectively.

baseline level of quality in the industry and the specific patterns of undertreatment, overtreatment, and overcharging that occur; and secondarily as providing suggestive though not dispositive evidence that moral hazard in the form of opportunistic behavior by mechanics is a primary determinant of the outcomes.

IV(i). *Evidence from the Field Experiment*

Table III shows how many times each legitimate defect was discovered by mechanics during the 40 experimental visits. The mode number of defects discovered was one, and in 21 of 40 visits (55 per cent), two or fewer defects were discovered. In only 4 visits (10 per cent) were a majority of the defects discovered. The missing tail light was found in 5 of 40 visits (13 per cent). The loose battery cable was corrected in 27 of 40 visits (68 per cent), though when (unnecessary) battery and starter motor replacement recommendations are included, which would have incidentally corrected the loose cable, the problem would be fixed in 31 of 40 visits (78 per cent).

Except for the loose battery cable, the defect requiring attention most urgently was the low coolant level, which was easily visible in the engine compartment and risks the life of the engine, especially during the hot summer months in which the visits occurred. This defect was discovered in only 11 of 40 visits (28 per cent), indicating serious undertreatment during most visits.[22] Failing to correct the low coolant level and the loose battery

[22] An alternative explanation for the poor coolant-level detection rate was intentional neglect with the aim of winning a large engine repair in the future. This possibility seems unlikely given the existence of many legitimate defects that could have been repaired immediately or in the future.

TABLE IV
UNNECESSARY REPAIRS RECOMMENDED DURING EXPERIMENTAL VISITS

	One-time	Repeat	Total
Replace starter motor	4	3	7
Replace battery	1	2	3
Replace radiator hoses	1	1	2
Replace thermostat	1	1	2
Replace water pump	2	0	2
Replace battery cable	0	1	1
Fuel injection cleaning	1	0	1
Replace fuel filter	0	1	1
Total	10	9	19

Note: The table reports incidences of unnecessary repairs recommended by mechanics during the 20 one-time and 20 repeat-business experimental visits.

cable represents the most serious neglect, and at least one of these problems was missed in 30 of 40 visits (75 per cent).[23]

Despite the existence of legitimate defects requiring attention, unnecessary repairs were recommended in many visits. Replacing a well-functioning starter motor or nearly-new battery, for example, was a regular prescription for the intermittent starting problem, occurring in 7 of 40 visits (18 per cent) and 3 of 40 visits (8 per cent), respectively, despite the presence of a visibly-loose battery cable. Table IV reports the full list of unnecessary repairs that were recommended.[24]

Since I did not consent to repairs, I cannot verify directly whether mechanics would have conducted repairs for which they charged. However, based on the types of repairs that were recommended, it appears likely the recommended repairs would have been conducted. The most frequent repairs were to the exhaust system, starter motor, spark plug wires and timing belt. In all of these cases, it would be clear to a moderately-informed motorist and to future mechanics if the repair were conducted.[25] In summary, the types of repairs that were recommended were fairly conspicuous, showing no indication that mechanics intended to charge for repairs they did not conduct. This result, combined with the limited amount of

[23] Also note that in addition to the low quality of most diagnoses, the repair recommendations were highly inconsistent across visits. In 22 of 40 visits, less than $50 in repairs were recommended, while in 12 visits, over $400 repairs were recommended, and in two visits, $1,398 and $1,849 in repairs were recommended.

[24] Conditional on repair type, prices were fairly uniform. For example, for the five visits in which starter-motor replacement was recommended and the price was listed separately, the prices were $190, $206, $235, $254, and $240. Variation in the prices for exhaust and belt service was larger, but mostly because different degrees of work were recommended. For example, mechanics recommended replacing different amounts of the exhaust system.

[25] For example, the exhaust system had conspicuous rust, which would make it difficult to claim the repair was conducted when it was not, while a timing belt that was not actually replaced would likely break in the future, causing serious engine damage, and might be traceable back to the mechanic.

TABLE V
UNNECESSARY REPAIRS RECOMMENDED DURING CANADIAN VISITS

	One-time	Repeat	Total
Replace spark plugs and/or wires	2	2	4
Replace starter motor	1	2	3
Replace battery	1	2	3
Replace distributor cap and rotor	1	2	3
Repair brakes	1	1	2
Replace fuel filter	1	1	2
Replace drive belt	0	2	2
Repair fuel injection	2	0	2
Replace PCV valve	0	1	1
Replace front struts	1	0	1
Replace ignition wires	1	0	1
Replace alternator	0	1	1
Total	11	14	25

Note: The table reports incidences of unnecessary repairs recommended by mechanics during the Canadian visits.

overcharging during the Canadian visits (described below), where researchers consented to all repairs, indicates that this practice is infrequent.

IV(ii). *Evidence from the Canadian Data*

APA mechanics estimated that a mechanic could easily diagnose and correct the loose battery cable in twenty minutes, and that the correction and general vehicle inspection should take no longer than sixty minutes. I allow an extra fifteen minutes to be conservative, and assume an hourly rate of $70, for an upper bound on a reasonable price for the visit of $88. Charges in excess of this amount when the battery cable was corrected but without additional work being performed are counted as overcharging. Charges in excess of this amount when additional work was conducted are counted as overtreatment.[26]

In 40 of 51 visits (78 per cent), the loose battery cable was corrected.[27] This rate is identical to the rate from the experimental data reported earlier. In 14 of 51 visits (27 per cent), overtreatment occurred, by an average amount of $244 conditional on overtreatment. In only 3 of 51 visits (6 per cent) did overcharging occur, and by an average amount of only $32 per incident. Table V contains the full list of unnecessary repairs that were conducted that cost over $50.[28]

[26] While these cutoff levels are arbitrary, the qualitative conclusions are robust to modifying these levels.

[27] The two visits in which the province of the license plate did not match the provided address are included for these calculations since the calculations do not involve one-time versus repeat-business appearance.

[28] There were two instances of sabotage of a vehicle part (documented by a hidden camera) to justify a repair, and repair prices were small in both cases.

Dividing the sum of overcharges across all 51 visits by the sum of total charges for the 51 visits reveals that only 2 per cent of total charges represent overcharging. The same calculation for overtreating reveals that a much larger 61 per cent of all charges represent completely unnecessary repairs. While the fraction of total repair costs that represents unnecessary repairs might depend on the base level of necessary repairs, the observed difference between overcharging and overtreating is large enough to indicate that overtreating is far more prevalent.

V. EVIDENCE ABOUT REPUTATION EFFECTS

V(i). *Evidence from the Field Experiment*

During the experiment, the average diagnosis fee for visits in which I represented one-time business was \$59.75, while the average price for visits in which I appeared as possible repeat business was \$37.70, for a difference of \$22.05 (p = 0.05). I next estimate the following model of one-time versus repeat-business on the diagnosis fee by ordinary least squares (OLS),

$$y_{ij} = \beta_0 + \beta_1 z_j + \mathrm{x}_i \beta_2 + w_j \beta_3 + \varepsilon_{ij}$$

where y_{ij} is the diagnosis fee charged by garage i during visit j; $z_j = 0$ indicates one-time business, and $z_j = 1$ indicates repeat business; x_i is a row vector of garage attributes including an indicator for whether the garage displayed an ASE sign, and garage size measured by number of active bays; w_j is a row vector of visit attributes including the time the vehicle was brought to the garage, and whether the Heat Index exceeded 90 F. that day; and ε_{ij} is a random error.[29,30]

The estimate from the base model in column (1) of Table VI shows that the diagnosis fee is \$22.01 (p = .04) lower when I represented possible repeat business. The estimate from the full specification in column (2) is \$26.17 (p = .02).

Next, I test whether inspection quality, measured as the number of legitimate defects discovered, is different for visits in which I represented possible repeat business. I estimate a Poisson maximum likelihood regression model with the following conditional mean function,

$$E\left[n_{ij} \middle| \cdot\right] = \exp\left\{\gamma_0 + \gamma_1 z_j + x_i \gamma_2 + w_j \gamma_3\right\}$$

[29] The Heat Index combines air temperature and relative humidity for an index reflecting how hot it actually feels. Since garage bays often lack air conditioning, the level of the Heat Index may affect mechanics' repair decisions.

[30] I also estimate a version of the model with fixed effects for each matched garage pair (20 in all). The estimate of the repeat-business effect is very similar and an F-test fails to reject that the fixed effects are jointly significant.

TABLE VI
MODELS OF DIAGNOSIS FEES

	(1) Exp. data	(2) Exp. data	(3) Canadian data	(4) Canadian + exp. data
Repeat business	−22.01**	−26.17**	−17.33**	−19.51***
	[10.31]	[10.76]	[7.14]	[6.09]
Arrival time		1.90		
		[3.39]		
ASE certification		14.70		
		[10.51]		
Garage size (no. of bays)		3.99		
		[4.52]		
Heat Index > 90F		23.87**		
		[11.66]		
Canadian visit				−8.84
				[6.09]
Constant	57.43***	14.31	46.09***	56.18***
	[7.29]	[41.23]	[5.41]	[5.40]
N	40	40	49	89
Obs. mean of dep. var.	46.43	46.43	36.14	40.87

Note: The dependent variable is diagnosis fee (in 2005 U.S. dollars). The models are estimated with OLS. Standard errors are reported in brackets. The models in columns (1) – (2), (3), and (4) are estimated with data from the experimental visits, Canadian visits, and both, respectively. *, **, and ***indicate significance at the 10, 5, and 1 per cent levels, respectively.

where n_{ij} is the number of legitimate defects discovered during the inspection, including the loose battery cable, and the regressors are defined as above. Column (2) in Table VII provides estimates from the full specification and shows that the possibility of repeat business increases the number of legitimate defects discovered by only 0.07, which is a negligible fraction of the mean number of discovered defects of 2.58.[31] Column (3) includes diagnosis fee as an explanatory variable and shows that diagnosis quality is increasing in diagnosis fee (though causality of course may act in both directions). The magnitude of the effect is relatively modest though significant at the 5 per cent level. Columns (4) and (5) provide alternate measures of diagnosis quality. The dependent variable (diagnosis quality) in column (4) is limited to the five defects that the APA deemed should be corrected immediately (loose battery cable, low coolant level, exhaust pipe leak, missing taillight, worn plugs and wires), and in column (5) is limited to the two most pressing defects (loose battery cable, low coolant level). Results are similar.

I next test for differences in repair prices between repeat and one-time business visits using the Mann-Whitney rank-sum test. This test fails by a wide margin to reject the hypothesis that mean repair prices for the two

[31] Estimates from a linear model, from a model that includes fixed effects for matched pairs, and the more flexible functional form of the negative binomial model all return very similar results. The model in column (3) uses a narrower measure of inspection quality, whether the battery cable, the low coolant level, or both are discovered during the inspection, and the results are also similar.

TABLE VII
MODELS OF DIAGNOSIS QUALITY

	(1)	(2)	(3)	(4)	(5)	(6)
	Exp. data	Exp. data	Exp. data	Exp. data: Alternate measure 1	Exp. data: Alternate measure 2	Canadian + exp. data: Battery
Repeat business	0.136	0.071	0.230	–0.057	0.035	–0.005
	[0.198]	[0.210]	[0.220]	[0.280]	[0.347]	[0.251]
Arrival time in hours		–0.082	–0.097	–0.085	0.038	
		[0.068]	[0.073]	[0.092]	[0.113]	
ASE certification		0.102	0.035	0.248	0.434	
		[0.206]	[0.211]	[0.271]	[0.335]	
Garage size (no. of bays)		0.148*	0.118	0.136	0.103	
		[0.083]	[0.086]	[0.111]	[0.136]	
Heat Index > 90F		0.255	0.087	0.168	0.075	
		[0.224]	[0.237]	[0.306]	[0.383]	
Diagnosis fee			0.007**			
			[0.003]			
Canadian visit						0.203
						[0.254]
Constant	0.875***	1.265	1.164	0.790	–0.982	–0.428*
	[0.144]	[0.807]	[0.860]	[1.088]	[1.382]	[0.233]
N	40	40	40	40	40	89
Obs. mean of dep. var.	2.575	2.575	2.575	1.450	0.950	0.730

Note: The dependent variable in columns (1)–(3) is diagnosis quality measured as the number of legitimate defects discovered (out of ten), in column (4), as the number of defects the APA mechanics deemed important to correct immediately (out of five), in column (5), as the two most pressing defects, which are the loose battery cable and the low coolant (out of two), and in column (6), as whether the battery cable was corrected, which is the common, primary defect in both the experimental and Canadian visits. The models in columns (1)–(5), and column (6) are estimated with the experimental and Canadian data respectively. The models are estimated as Poisson regression models and estimates are reported as marginal effects evaluated at the means of the other explanatory variables. Standard errors are reported in brackets. *, **, and ***indicate significance at the 10, 5, and 1 per cent levels, respectively.

TABLE VIII
MODELS OF REPAIRS

	(1) Exp. data	(2) Exp. data	(3) Canadian data	(4) Canadian + exp. data
Repeat business	–0.050	–0.058	0.041	0.001
	[0.156]	[0.176]	[0.120]	[0.100]
Arrival time in hours		–0.114*		
		[0.062]		
ASE certification		–0.423***		
		[0.152]		
Garage size (no. of bays)		0.062		
		[0.073]		
Heat Index > 90F		–0.113		
		[0.188]		
Canadian visit				–0.201**
				[0.099]
N	40	40	49	89
Obs. mean of dep. var.	0.425	0.425	0.224	0.315

Note: The dependent variable is an indicator for whether at least one repair over $50 was recommended (experimental visits) or conducted (Canadian visits). The models are estimated as probit models and estimates are reported as marginal effects evaluated at the means of the other explanatory variables. Standard errors are reported in brackets. The models in columns (1)–(2), (3), and (4) are estimated with data from the experimental visits, Canadian visits, and both, respectively. *, **, and ***indicate significance at the 10, 5, and 1 per cent levels, respectively.

groups are the same (p = .76).[32] I next estimate the index model of whether over $50 in repairs were recommended,

$$Pr\left[r_{ij}=1\right]=\Phi\left[\alpha_0+\alpha_1 z_j+x_i\alpha_2+w_j\alpha_3\right]$$

where r_{ij} indicates whether repairs were recommended, Φ is the evaluation of the standard normal CDF, which implies the usual probit specification, and the regressors are defined as before. Column (2) in Table VIII contains estimates from the full specification. The estimate of the repeat business effect is small in magnitude and not statistically distinguishable from zero, providing no evidence that reputational concerns affect the probability of repairs. Note however that the modest sample size limits the power of this test. Column (4) contains an estimate of the repeat-business effect using pooled Canadian and experimental data, providing moderately better precision, and the estimate mirrors those from the individual sets of visits.

V(ii). *Evidence from the Canadian Data*

I first test whether the possibility of repeat business affects the diagnosis fee charged by mechanics. For Canadian visits in which the researcher repre-

[32] There are insufficient data to test whether the mean repair price for repeat-business visits was lower than one-time visits conditional on repair type, though no obvious differences are apparent.

TABLE IX
DIAGNOSIS FEES FOR VISITS WITH AND WITHOUT REPAIRS

	Experiment		Canadian	
	Repairs	No repairs	Repairs	No repairs
One-time	39.68	47.92	54.26	61.30
Repeat	31.83	26.65	47.03	25.92
Difference	7.84	21.27**	7.23	35.38**
	[14.47]	[7.83]	[15.41]	[13.71]
N	20	20	17	32

Note: The table reports mean diagnosis fees for the experimental visits and Canadian visits (in 2005 U.S. dollars), for visits in which repairs are recommended/conducted versus not. One-time indicates the undercover researchers appeared as one-time business. Repeat indicates they appeared as possible repeat business. The standard error of the difference between one-time versus repeat business visits are reported in brackets. **indicates significance at the 5 per cent level.

sented possible repeat business, the average diagnosis fee was $28.76, while for visits in which the researcher represented one-time business, the average fee was $46.09. The difference is $17.33 and statistically significantly different from zero ($p = .02$).[33,34] Mechanics appear to discount the diagnosis fee substantially for motorists who represent possible repeat business.[35]

As a function of the respective mean diagnosis prices, these estimates are nearly identical to the estimate from the experimental visits; both sets of results are provided for comparison in Table IX. Column (4) of Table VI provides a more precise estimate of the repeat-business effect from the pooled Canadian and field experiment data, showing that the diagnosis fee was $19.51 lower for repeat business visits ($p < .01$).

The repeat-business effect does not carry over to repairs. In 7 of the 29 repeat-business visits (24 per cent), the mechanic conducted at least $50 in repairs, while in 4 of the 20 one-time-business visits (20 per cent), the mechanic conducted at least $50 in repairs (recall that no repairs were required).[36] While the rate of overtreatment under repeat business is higher than under one-time business, a z-test does not reject the hypothesis that repair probabilities for the two groups are the same ($p = .73$). I also test whether the repair prices between one-time versus repeat-business visits differ. Since the distribution over repair prices is highly skewed, I use the

[33] Estimating a model with OLS with diagnosis fee as the dependent variable on dummy variables for garage types (franchise, chain, dealer, independent) and one-time versus repeat business gives a very similar estimate of the repeat business effect. The results were also unaffected by including a cost-of-living measure for the city of the garage, and an indicator for whether the loose battery cable was corrected.

[34] Diagnosis fees for two visits are excluded here and in Table VI because receipts did not break out these charges.

[35] The mean diagnosis fees for chains and franchises are $37.93 and $31.06 respectively, and the difference is not statistically different from zero ($p = .33$). Only one dealer and four independent garages are in the Canadian data, precluding analysis of differences between these garage types.

[36] The $50 cutoff is arbitrary, but the results are qualitatively robust to changes in this level.

Mann-Whitney rank-sum test, and again find no statistical difference between the groups. Finally, I estimate the model of inspection quality from the previous subsection but with the pooled Canadian and experimental data, using as the measure of inspection quality whether the loose battery cable was corrected (the common defect between the two sets of visits). The results are in column (6) of Table VII and confirm the absence of a repeat-business effect.

VI. DISCUSSION OF RESULTS

The results indicate that overtreating in the auto repair market is common while overcharging is not. Previous theoretical work (e.g., Alger and Salanie [2006], Dulleck and Kerschbamer [2006]) highlights the role of the verifiability condition—that buyers can observe which repairs were conducted after the transaction (though not whether they were necessary)—in leading sellers to conduct unnecessary service instead of simply charging for service they do not provide. The current results are consistent with the predicted effect of this condition.

The prevalence of undertreatment highlights the importance of the failure of the liability condition (e.g., Dulleck and Kerschbamer [2006])—that sellers are not penalized for failing to correct legitimate defects, either because the buyer does not learn of this negligence or the negligence is not legally or otherwise penalized—in facilitating undertreatment. The prevalence of undertreatment also indicates an important role for effort during the diagnosis process, as highlighted in Pesendorfer and Wolinsky [2003]. Several defects, such as the missing taillight, were trivial to correct once discovered, but finding them requires modest but still costly effort.

The failure of reputation to significantly limit agency problems could be a consequence of motorists being unable to evaluate service quality after the transaction, which is a common assumption in the reputation literature. The finding of a large repeat-business effect for the diagnosis fee, on the other hand, may reflect motorists' ability to evaluate this fee after service. For example, in 7 of 40 visits, the inspection was free, which motorists could evaluate as favorable (notably 6 of which were repeat-business visits), while in two visits, the inspection cost over $100 which would be less favorable (both were one-time visits).

In support of this explanation, note that if motorists' ability to evaluate the diagnosis fee is indeed important, the repeat-business discount should be less pronounced for visits with repairs since motorists may be unable to evaluate whether a lower diagnostic fee is simply offset with more repairs, and hence the value to the mechanic of charging a lower diagnosis fee is more limited. This pattern is indeed evident. Table IX reports the mean diagnosis fees for one-time versus repeat-business visits depending on whether repairs are recommended. In the experimental visits, for visits with

repairs, the fee was $7.84 lower for repeat visits versus one-time visits ($p = .64$). However, for visits without repairs, the fee was $21.27 lower ($p = .02$). For the Canadian visits with repairs, the fee was $7.23 lower for repeat-business visits versus one-time visits ($p = .60$). However, for visits without repairs, the fee was $35.38 lower ($p = .01$). As noted earlier, the finding in List [2006b] that reputation and buyer learning are complements indicates the importance of this monitoring. The current results are consistent this finding.

VI(i). *Alternative Explanations*

It is worth discussing an alternate explanation for the observed low quality service and limited repeat-business effect, which is widespread incompetence among mechanics. Under this scenario, when most mechanics face an intermittent starting failure, they simply do not have the basic skills to diagnose the cause. Furthermore, these mechanics do not know basic aspects of a rudimentary vehicle inspection, which are to check the lights and fluids. Under this scenario, testing for a difference in performance between one-time and repeat visits would reveal no repeat-business effect since even well-intentioned mechanics cannot identify problems correctly.

I believe this explanation is unlikely for several reasons. First, as the photograph of the engine compartment in Appendix A shows, the loose battery cable was highly visible in the front of the engine compartment; diagnosing this problem required little more than a visual inspection of the engine compartment. Second, the mechanics working with the Canadian group indicated that these defects should be identifiable to even the least experienced mechanics. Many of the defects were chosen because they are among the simplest to diagnose that mechanics ever face. Third, I examine whether service quality varies with two garage attributes that may reflect mechanic skill. The idea is that if the poor outcomes are due to low skill and not incentives, then service quality should improve with measures of skills.

I first test whether establishments that have been in operation for a longer period of time, which may reflect more experience on average (in particular for the independent garages that are the focus of the experiment, since the owner is involved in most decisions), provide higher-quality service. I was able to obtain establishment age for 32 garages based on data from Dun & Bradstreet, and divided the garages into those in operation for 0 to 9 years, 10 to 19 years, and at least 20 years (11, 8, and 13 garages, respectively). I find no relationship between age and performance: Garages with 0 to 9, 10 to 19, and at least 20 year-old garages found the battery cable problem 73, 75, and 46 per cent of the time, and diagnosed 2.2, 3.3, and 2.0 legitimate defects on average, respectively.

I next test whether outcomes are better for garages with ASE-certified mechanics. ASE certification requires passing at least one written auto repair test, which is administered by ACT (formerly American College Testing, the organization that administers the ACT college entrance exam), focusing on a particular car system, and documenting two years of relevant work experience. Sixty-five per cent of the 17 ASE garages corrected the loose battery cable problems, while 65 per cent of the 23 non-ASE garages also corrected the problem, indicating no difference for the certified group. While the garages with ASE-certified mechanics discovered 2.88 legitimate defects, which is modestly more than the 2.35 defects discovered by non-ASE-certified garages, the difference represents only a small fraction of the standard deviation of the number of legitimate defects discovered of 1.68, indicating that this measure of mechanic skill explains only a modest fraction of this variation.

There are several alternative ways in which the repeat-business procedure could generate less treatment and lower prices independent of a reputation mechanism that are less compelling but worth acknowledging. First, it is conceivable that mechanics would view a round-trip to Montreal without moving boxes as less demanding than a one-way trip to Chicago with boxes, which may cause them to conduct fewer repairs proactively. Second, mechanics may postpone recommending repairs to possible repeat customers since they expect to be able to conduct these repairs in subsequent visits. Third, mechanics may expect possible repeat customers who are new to the neighborhood to search more thoroughly for a mechanic versus regular repeat customers, and hence may offer fewer repairs and lower prices only in the initial visits but not on subsequent repeat visits. None of these factors, however, are relevant for whether a mechanic should address most of the legitimate defects (for example, the loose battery cable or low coolant level), and all of these factors would cause the repeat-business procedure to generate even fewer repairs and lower prices than a repeat-business mechanism alone. Since I do not find a repeat-business effect, these alternative factors do not appear to be important.

VII. CONCLUDING REMARKS

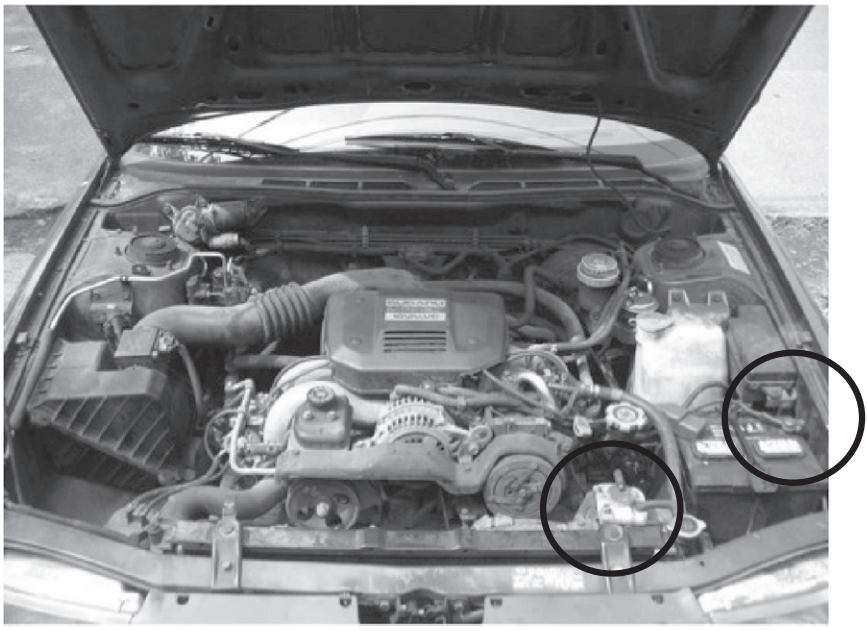
The results of this study are consistent with large agency problems in the market for auto repairs, and that mechanic are no more likely to provide efficient service when a motorist represents repeat versus one-time business. While it is beyond the scope of this study to analyze specific remedies, I mention several possibilities now. One involves the public posting of outcomes from third-party evaluations of service quality. This type of program has been moderately effective in at least one other setting. The Los Angeles County Department of Health requires restaurants to publicly post the outcomes of their quarterly hygiene inspections (New York City recently

instated a similar policy). Jin and Leslie [2003] find modest increases in hygiene scores and reductions in hospitalizations for food-borne illness as a result of this policy.

Another possibility is businesses' providing diagnosis-only inspections, which would reduce mechanics' incentives to recommend too much treatment, though would still require costly inspections. Finally, the public posting of buyers' experiences, so that a history of sellers' actions is observable, could help. Even in settings where individual buyers cannot evaluate the quality of service they receive, a history of the mechanic's actions may reveal clear patterns of behavior. The publication, *Consumers' Checkbook*, and websites such as *Yelp* and RepairPal are emerging platforms for disseminating such information.

APPENDIX A
PHOTOGRAPHS OF THE TEST VEHICLE

Note: The photograph shows the experimental test vehicle, prepared for a repeat-business procedure with moving material and boxes in the car, with the location of the missing tail light circled. Photo by author.



Note: The photograph shows the engine compartment of the experimental test vehicle, with the location of the connection of the ground battery cable to the battery, which was made conspicuously loose, in the right circle, and the coolant overflow tank, which was emptied, in the left circle. Photo by author.

APPENDIX B
FIELD EXPERIMENT SCRIPT

Calling for Appointment. Hi, I'm wondering if I can arrange a time when I can bring in my car for service? We bought the car recently and it hasn't started a few times—can you check that out? We also want to get a thorough inspection to see if any other work needs to be done. When's the soonest I can bring it in?

If Asked to Keep the Car Overnight. My wife needs the car in the evenings. Can I just bring it in on a day when you'll have time to look at it? I can drop it off early.

Script for Low-Reputation Procedure. My wife and I are moving to Chicago in two weeks and we're taking the car with us—we just wanted to have some things looked at before the trip. I scheduled an appointment for this morning. We bought the car recently and we should have had it looked at before we bought it, but we didn't. It hasn't started a few times—can you check that out? We'd also like a thorough inspection to let us know if any other work needs to be done. Will you give me a call to let me know how things are going? Roughly how long do you think that'll take?

If Asked for an Address. Our new address in Chicago is 203 Water Street. We're staying with friends down the street for now, on [local street].

Script for High-Reputation Procedure. I'm moving in just down the street so I figured I'd come check you guys out. I scheduled an appointment for this morning. My wife and I are taking a trip to Montreal in two weeks—and we're taking the car with us. We just wanted to have some things looked at before the trip. We bought the car recently and we should have had it looked at before we bought it, but we didn't. It hasn't started a few times—can you check that out? We'd also like a thorough inspection to let us know if any other work needs to be done. I'm going to run home for a few minutes. Will you give me a call to let me know how things are going? How long do you think that'll take?

If Asked about the Car's Service History. We haven't had anything done since we got the car a couple of months ago. Just an oil change.

If Asked about the Timing Belt. I'm not sure—I don't know much about it.

If Asked whether the Previous Owner could be Contacted about the Car's Service History. It's someone my dad knew. It'd be hard to contact him.

If Recommendations are Ambiguous. What work do you think we should do?

Declining Service. Let me talk to my wife—she drives the car most of the time.

REFERENCES

Alger, I. and Salanie, F., 2006, 'A Theory of Fraud and Overtreatment in Experts Markets,' *Journal of Economics and Management Strategy*, 15(4), pp. 853–881.

Andreoni, J. and Croson, R., 2008, 'Partners Versus Strangers: Random Rematching in Public Goods Experiments,' *Handbook of Experimental Economics Results*, 1, pp. 776–783.

Balafoutas, L.; Beck, A.; Kerschbamer, R. and Sutter, M., 2011, 'What Drives Taxi Drivers? A Field Experiment on Fraud in a Market for Credence Goods,' unpublished manuscript, *http://ftp.iza.org/dp5700.pdf*

Bar-Isaac, H. and Tadelis, S., 2008, 'Seller Reputation,' *Foundations and Trends in Microeconomics*, 4(4), pp. 273–351.

Banerjee, A. and Duflo, E., 2000, 'Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry,' *Quarterly Journal of Economics*, 115(3), pp. 989–1017.

Beck, A.; Kerschbamer, R.; Qiu, J. and Sutter, M., 2009, 'Car Mechanics in the Lab—Investigating the Behavior of Real Experts on Experimental Markets for Credence Goods,' unpublished manuscript, *http://eeecon.uibk.ac.cet/wopecz/repec/inn/wpapec/2009-27.pdf*

Bolton, G.; Katok, E. and Ockenfels, A., 2004, 'How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation,' *Management Science*, 50(11), pp. 1587–1602.

Darby, M. and Karni, E., 1973, 'Free Competition and the Optimal Amount of Fraud,' *Journal of Law and Economics*, 16(1), pp. 67–88.

Dulleck, U. and Kerschbamer, R., 2006, 'On Doctors, Mechanics and Computer Specialists: The Economics of Credence Goods,' *Journal of Economic Literature*, 44(1), pp. 5–42.

Dulleck, U.; Kerschbamer, R. and Sutter, M., 2011, 'The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation and Competition,' *American Economic Review*, 101, pp. 526–555.
Ellison, G. and Chevalier, J., 1999, 'Career Concerns of Mutual Fund Managers,' *Quarterly Journal of Economics*, 114(2), pp. 389–432.
Ely, J. and Valimaki, J., 2003, 'Bad Reputation,' *Quarterly Journal of Economics*, 118(3), pp. 785–814.
Ely, J.; Fudenberg, D. and Levine, D., 2008, 'When is Reputation Bad?' *Games and Economic Behavior*, 63(2), pp. 498–526.
Emons, W., 1997, 'Credence Goods and Fraudulent Experts,' *RAND Journal of Economics*, 28(1), pp. 107–119.
Emons, W., 2001, 'Credence Goods Monopolists,' *International Journal of Industrial Organization*, 19(3–4), pp. 375–389.
Fong, Y., 2005, 'When Do Experts Cheat and Whom Do They Target?' *RAND Journal of Economics*, 36(1), pp. 113–130.
Gaynor, M. and Vogt, W., 2000, 'Antitrust and Competition in Health Care Markets,' in A.J. Cuyler and J.P. Newhouse, eds., *Handbook of Health Economics, Vol. 1*, Elsevier, chapter 20.
Grosskopf, B. and Sarin, R., 2010, 'Is Reputation Good or Bad? An Experiment,' *American Economic Review*, 100(5), pp. 2187–2204.
Hubbard, T., 1998, 'An Empirical Examination of Moral Hazard in the Vehicle Inspection Market,' *RAND Journal of Economics*, 29(2), pp. 406–426.
Hubbard, T., 2002, 'How Do Consumers Motivate Experts? Reputational Incentives in an Auto Repair Market,' *Journal of Law and Economics*, 45(2), pp. 437–468.
Iizuka, T., 2007, 'Experts' Agency Problems: Evidence From the Prescription Drug Market in Japan,' *RAND Journal of Economics*, 38(3), pp. 844–862.
Jin, G. and Leslie, P., 2003, 'The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards,' *Quarterly Journal of Economics*, 118(2), pp. 409–451.
Jin, G. and Leslie, P., 2009, 'Reputation Incentives for Restaurant Hygiene,' *American Economic Journal: Microeconomics*, 1(1), pp. 237–267.
Klein, B. and Leffler, K., 1981, 'The Role of Market Forces in Assuring Contractual Performance,' *Journal of Political Economy*, 89(4), pp. 615–641.
Kreps, D.; Milgrom, P.; Roberts, J. and Wilson, R., 1982, 'Rational Cooperation in the Finitely Repeated Prisoners' Dilemma,' *Journal of Economic Theory*, 27, pp. 245–252.
Kreps, D. and Wilson, R., 1982, 'Reputation and Imperfect Information,' *Journal of Economic Theory*, 27(2), pp. 253–279.
Levitt, S. and Syverson, C., 2008, 'Market Distortions when Agents are Better Informed: The Value of Information in Real Estate Transactions,' *Review of Economics and Statistics*, 90 (4), pp. 599–611.
List, J., 2006a, 'Field Experiments: A Bridge Between Lab and Naturally Occurring Data,' *Advances in Economic Analysis and Policy*, 6(2).
List, J., 2006b, 'The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions,' *Journal of Political Economy*, 114(1), pp. 1–37.
Mailath, George and Samuelson, L., 2001, 'Who Wants a Good Reputation?' *Review of Economic Studies*, 68, pp. 415–441.
McGuire, T., 2000, 'Physician Agency,' in A.J. Cuyler and J.P. Newhouse, eds., *Handbook of Health Economics, Vol. 1*, Elsevier, chapter 9.
Milgrom, P. and Roberts, J., 1982, 'Predation, Reputation, and Entry Deterrence,' *Journal of Economic Theory*, 27(2), pp. 280–294.

Palfrey, T. and Prisbrey, J., 1996, 'Altruism, Reputation and Noise in Linear Public Goods Experiments,' *Journal of Public Economics*, 61(3), pp. 409–427.
Park, I., 2005, 'Cheap Talk Referrals of Differentiated Experts in Repeated Relationships,' *RAND Journal of Economics*, 36, pp. 391–411.
Pesendorfer, W. and Wolinsky, A., 2003, 'Second Opinions and Price Competition: Inefficiency in the Market for Expert Advice,' *Review of Economic Studies*, 70(2), pp. 417–437.
Pitchik, C. and Schotter, A., 1987, 'Honesty in a Model of Strategic Information Transmission,' *American Economic Review*, 77(5), pp. 1032–1036.
Schneider, H., 2010, 'Moral Hazard in Leasing Contracts: Evidence from the New York City Taxi Industry,' *Journal of Law and Economics*, 53, pp. 783–805.
Shapiro, C., 1983, 'Premiums for High Quality Products as Returns to Reputation,' *Quarterly Journal of Economics*, 98(4), pp. 659–674.
Taylor, C., 1995, 'The Economics of Breakdowns, Checkups and Cures,' *Journal of Political Economy*, 103(1), pp. 53–74.
Wolinsky, A., 1993, 'Competition in a Market for Informed Experts' Services,' *RAND Journal of Economics*, 24(3), pp. 380–398.