Technological Shocks and Algorithmic Decision Aids in Credence Goods Markets

Alexander Erlei* Lukas Meub†

Abstract

In credence goods markets such as health care or repair services, consumers rely on experts with superior information to adequately diagnose and treat them. Experts, however, are constrained in their diagnostic abilities, which hurts market efficiency and consumer welfare. Technological breakthroughs that substitute or complement expert judgments have the potential to alleviate consumer mistreatment. This article studies how competitive experts adopt novel diagnostic technologies when skills are heterogeneously distributed and obfuscated to consumers. We differentiate between novel technologies that increase expert abilities, and algorithmic decision aids that complement expert judgments, but do not affect an expert's personal diagnostic precision. We show that high-ability experts may be incentivized to forego the decision aid in order to escape a pooling equilibrium by differentiating themselves from low-ability experts. Results from an online experiment support our hypothesis, showing that high-ability experts are significantly less likely than low-ability experts to invest into an algorithmic decision aid. Furthermore, we document pervasive under-investments, and no effect on expert honesty.

JEL codes. C91, C92, C72, D82

Keywords. Credence goods, expert ability, diagnostic uncertainty, technology, decision aid, experiment

*Corresponding author. Chair for Economic Policy and SME Research, Georg-August University of Goettingen. alexander.erlei@wiwi.uni-goettingen.de

†Institute for Small Business Economics (ifh), Georg-August-University Goettingen.

We gratefully acknowledge financial support by the Federal Ministry of Education and Research, project "Handwerk mit Zukunft (HaMiZu)", grant number 02K20D001.

Credence goods markets are characterized by severe information asymmetries between experts and consumers. This often translates into market inefficiencies, with experts either resorting to insufficient treatments, or using their information advantage to overcharge consumers, e.g. for unnecessary services. Due to the ubiquity of expert services throughout important economic decision domains such as medicine, law, financial advice or accounting, these patterns are highly consequential, and very costly for both consumers and the overall economy. The theoretical literature usually attributes these adverse outcomes to expert fraud. That is, given the right incentives, experts would behave "honestly" and treat consumers perfectly. However, in reality, a second factor that can severely inhibit the provision of high-quality services is the expert's *diagnostic ability*. While this is obvious in a one-shot environment, the information asymmetry in credence goods markets generally precludes many consumers from adequately judging the quality of a performed service ex post. This reduces their ability to effectively discipline experts, who in turn have little incentive to invest in their abilities. For example, Zhi et al. (2013) document an under-utilization rate of 44.8% for 46 of the most commonly used diagnostic rests between 1997 and 2012. Simultaneously, around 30% of health expenditures in the US are used for non-beneficial interventions (Brody, 2010). Several studies show the detrimental effect of insufficient skills on misdiagnoses as well as faulty prescription of aggressive medication such as antibiotics (Chan et al., 2022; Xue et al., 2019; Currie and MacLeod, 2017). Field experiments in the car repair industry point to under-treatment rates of around 75% (Schneider, 2012), as well as a negative correlation between competency and low-quality services (Rasch and Waibel, 2018). In auditing, heterogeneity in expert ability forces consumers to pay high premiums for skilled service providers (Aobdia et al., 2021).

Possibly the most promising solution to these entrenched quality issues are technological breakthroughs, specifically in algorithmic decision systems. Even today, prediction models complement or substitute human judgments in many professional fields, including legal search and bail decisions (Kleinberg et al., 2018), clinical diagnosis, medical scans (Topol, 2019) or financial advice (Brenner and Meyll, 2020). Large language models such as GPT-4 are increasingly able to assist lawyers in discovering relevant information, generate and check legal documents, or predict court outcomes (Alarie et al., 2018). According to a study by the Michigan Legal Help Program, clients with access to a legal advice software were equally successful as clients with an attorney when filing for divorce (Sandefur, 2020). Machine learning models have been shown to substantially exceed physicians in predicting heart attacks by avoiding faulty human heuristics (Mullainathan and Obermeyer, 2022), and provide improved cancer screening (Daysal et al., 2022). However, in order to actually realize these potentially large welfare gains, experts need to adopt algorithmic decision systems at scale. If that happens, or how long it might take, is an open, but very consequential question. Humans appear to have persistent biases against algorithmic decisions, especially once they (inevitably) observe errors (Dietvorst et al., 2015). There are also reputational concerns. Relying on an algorithmic decision system precludes the expert from sending a competence signal, and may well indicate the opposite (Arkes et al., 2007; Dai and Singh, 2020).

In this paper, we experimentally investigate the effects of technological shocks on verifiable credence goods markets with diagnostic uncertainty and obfuscated, heterogeneous (high vs. low) expert abilities. Our experimental setup differentiates between two different phases. In a first phase, expert abilities are exogenous, and cannot be improved through investment. Contrary to prior experiments on credence goods, experts diagnose consumers problems by completing a short prediction task. High-ability experts are able to utilize more input variables than low-ability experts. All diagnoses are uncertain. Consumers can never observe an expert's ability type. This traps high-ability experts in a pooling equilibrium with low-ability experts that precludes them from exploiting their skill advantage. The second phase simulates an external technological shock. Depending on treatment, experts receive the option to increase their diagnostic precision by either (1) investing into their own abilities (*Skill*) or (2) investing into an algorithmic decision aid (*Algorithm*). This allows us to capture some distinct features of algorithmic decision aids, and compare them with a baseline where technological improvements are perfectly predictable and quantifiable. Technological investments are always transferred onto consumers via prices, representing e.g. many medical situations in which certain tests are not covered by insurance. We hypothesize that high-ability experts may be incentivized to forego the algorithmic decision aid, but not the skill investments, in order to signal consumers their personal ability type. The intuition is that with decision aids, an expert's personal ability type can still play a role in the consumer's choice process. Decision aids carry uncertainty, because consumers cannot perfectly predict whether an expert is gonna (correctly) utilize the system's information. Consumers also know that high-ability experts derive relatively less benefits from the decision aid, and may therefore rather compete through lower prices. Low-ability experts, on the other hand, cannot afford to imitate non-investing high-ability experts in the presence of Bayesian consumers. This allows high-ability experts to escape the prior equilibrium, and then exploit their skill advantage. To the best of our knowledge, this is the first paper to analyze obfuscated expert abilities as a potential barrier to technological adoption in professional services. Second, our experiment also empirically analyzes the efficiency of credence goods markets when experts are heterogeneous and diagnosis is uncertain, as well as the concurrent efficacy of expert investments after a technological shock.

Experimental results confirm the importance of obfuscated expert ability types for the costly adoption of algorithmic decision aids. In line with our predictions, there are no differences in investment behavior between ability types for *Skill*, whereas high-ability experts are much less likely to purchase the algorithmic decision aid than low-ability experts. Furthermore, high-ability experts who forego the algorithm exhibit different price-setting patterns from those who decide to invest, suggesting signalling behavior. Overall, the market is remarkably efficient, with almost full consumer participation despite common under- and overtreatment. Experts generally under-invest, which may be partially driven by consumers not rewarding beneficial investment patterns. Investments tend to shift the expert's price menus towards more self-serving ones that reward fraudulent undertreatment.

Our paper relates to the literature on credence goods markets with diagnostic uncertainty. Most articles are theoretical, showing how price competition may negatively affect diagnostic precision when experts need to exert effort (Pesendorfer and Wolinsky, 2003), that separating diagnosis and treatment can achieve first-best outcomes when consumer evaluations are subjective (Bester and Dahm, 2018), or that fraud-sensitive penalties can discipline experts whose diagnoses require proper incentives (Chen et al., 2022). In recent empirical work, Balafoutas et al. (2020) show in a laboratory experiment that diagnostic uncertainty decreases efficient service provision and consumer market entry. Insurances reduce investments into diagnostic precision, while pro-social experts invest more. There is also evidence that diagnostic uncertainty does not affect expert dishonesty or consumer trust, while reputation mechanisms remain beneficial for overall market efficiency (Tracy et al., 2023).

A second relevant string is the literature on heterogeneous expert abilities on credence goods markets. Prior theoretical work shows that efficient equilibria are always possible with heterogeneous, obfuscated expert abilities, but inefficient ones also exist (Liu et al., 2020). Incentives to invest into expert abilities may break down in the presence of discounters (Dulleck and Kerschbamer, 2009), and heterogeneous cost functions generally prevent first-best solutions (Hilger, 2016; Frankel and Schwarz, 2014). Schneider and Bizer (2017a) find that in the presence of low-ability experts who always misdiagnose consumers without exerting effort, a sufficient number of high-ability experts allows for a second-best equilibrium without policy intervention. Empirically, they show with a laboratory experiment that high-ability experts invest less than theoretically predicted, that the market is more efficient than predicted, and that price competition hurts overall welfare (Schneider and Bizer, 2017b).

Third and finally, this article contributes to the literature on decision aids in professional contexts. Especially in the medical literature, it is a widespread sentiment that experts who strongly rely on complementary decision expertise e.g., through laboratory testing, tend to be less competent (Schroeder et al., 1974; Daniels and Schroeder, 1977; Yeh, 2014; Hall et al., 2019; Miyakis et al., 2006; Doi et al., 2021; Groopman and Prichard, 2007). There is evidence that both patients and peers derogate the diagnostic ability of physicians who rely on diagnostic aids (Cruickshank, 1985; Arkes et al., 2007; Promberger and Baron, 2006; Shaffer et al., 2013; Wolf, 2014). Patients have formulated preferences for physicians using intuition-based diagnoses versus statistical models (Eastwood et al., 2012) and general aversions towards statistical methods (Dawes et al., 1989). While we do not measure consumers preferences or attitudes towards decision aids, we do study how experts may strategically avoid algorithmic support systems in order to influence consumer beliefs about the expert's ability type. Rather than projecting a general distrust of technology, we thereby explore another possibility why the adoption of algorithmic systems may lag behind the optimal investment path. Skilled experts use the appearance of a novel diagnostic tool to reduce consumer uncertainty about their ability and thereby differentiate themselves from low-ability experts.

1 The Market

We study a credence goods market with competing experts who are not liable for offering insufficient treatments. Consumers know that they have a big problem with probability h or a small problem with probability (1-h). Once consumers approach an expert, they are committed to receive the recommended treatment under the offered price, meaning there is no separation of diagnosis and treatment, and consumers cannot consult other information sources. Afterwards, consumers can verify the treatment and observe their payoff. If the problem is solved, consumers receive a payoff of v, if it is not solved, they earn nothing. We assume that indifferent consumers choose to visit the expert. Similarly, experts who are indifferent between cheating and treating the consumer honestly always choose the correct treatment. There are two commonly known expert types, high-ability (H, with probability γ) and low-ability (L, with probability $1-\gamma$) experts, who differ in the probability to perform a costless correct diagnosis: $z \in [0.75, 1]$ vs. $q \in [0.5, 0.75]$. Let $i \in \{H, L\}$ and $k \in \{z, q\}$. Experts always charge for the implemented treatment (verifiability), and a high quality treatment HQT (low quality treatment LQT) treatment induces costs of $\bar{c}$ ($\underline{c}$), with $\bar{c} > \underline{c}$. The HQT solves both problems, the LQT only solves the small problem. Expert type ability probabilities are common knowledge. Consumers therefore consider two possible types of misdiagnosis: diagnosing a major problem when they actually have a minor problem, and diagnosing a minor problem in case of a major problem. Consumers cannot observe an expert's ability type. They can, however, identify each individual expert, allowing for reputation. It is by assumption always profitable for the expert to treat the consumer because $v > \bar{c}$. Experts choose a price menu $\mathbf{P} = (\bar{p}, \underline{p})$ after learning about their type, with $\bar{p}$ (major treatment) $\geq \underline{p}$.

Investments. After n rounds, experts can choose to increase their diagnostic precision k for a fixed fee d. Investments automatically increase $\mathbf{P}$ by d. We assume that the low-ability expert matches the high-ability expert's skill by investing in the new technology. Thus, the low-ability expert improves their diagnostic precision by an additional $z - q$ for any chosen precision level. There are two reasons for that. First, in many cases, low-ability experts presumably make mistakes that are comparatively easy to eliminate. This also implies low costs. Second, investing specifically in tangible new technology like an algorithmic decision system does not only complement the expert's diagnosis, but often serves as an imperfect substitute for it Doyle Jr et al. (2010); Dai and Singh (2020). Assuming that the low-ability expert's mistakes are relatively easy to fix, we'd expect algorithmic decision aids to mostly eliminate them.[1]

[1]Even if the assumption that low-ability experts completely catch up to high-ability experts might be too strong for many real-life cases, it is almost certainly true that the marginal benefits of algorithmic systems will on average be larger for low-ability experts.

2 Experimental Design

We conduct two treatments of a pre-registered between-subject (*Skill* vs. *Algorithm*) online group experiment where three experts compete over three clients. The experimental parameters are fixed across treatments. Consumers have a big problem with probability $h = 0.4$, receive $v = 150$ if their problem is solved, and earn $\sigma = 15$ when choosing the outside option. The cost of providing the LQT is $\underline{c} = 20$, the cost for the HQT is $\bar{c} = 60$. Experts choose between three price vectors. The price for the HQT is $\bar{p} = 100$ and fixed. For $\mathbf{P}^m$, we set $\underline{p} = 40$; for $\mathbf{P}^e$, we set $\underline{p} = 60$; and for $\mathbf{P}^s$, we set $\underline{p} = 80$. Experts always have a markup of 40 for the HQT, and markups of 20/40/60 for the LQT when choosing $\mathbf{P}^m/\mathbf{P}^e/\mathbf{P}^s$. In each group, there are 2 low-ability experts and 1 high-ability expert. Low-ability (High-ability) experts receive the correct signal with a probability of 50% (75%). Investing into additional diagnostic precision costs $d = 10$ and increases an expert's diagnostic precision to $k = 0.9$. Importantly, investments also automatically increase $\underline{p}$ and $\bar{p}$ by 10 Coins, meaning that consumers pay for improved diagnoses. We apply the following conversion rate: 100 Coins = 60 cents. Data from 300 participants (40% female) was collected on Amazon Mechanical Turk (MTurk) using oTree (Chen et al., 2016) and CloudResearch (Litman et al., 2017). All participants are vetted for quality by CloudResearch, completed at least 50 prior tasks on MTurk with a minimal approval rating of 90%, and reside in the US. The base payment was $4.50.

2.1 Procedure

Subjects play 25 rounds in groups of six. The first 10 rounds are constant across treatments and always follow the same sequence. Experts choose their price vector, and then diagnose all three consumers by completing a short prediction task. Here, high-ability experts use 3 out of 5 possible input factors, and low-ability experts use 2 input factors to identify the problem. Input numbers are pre-filled for each expert, who automatically make a prediction by clicking on a "diagnose" button. Then, experts receive a diagnostic signal that depends on their ability type. Finally, experts choose for each consumer whether they want to implement the HQT or the LQT, and proceed to a summary screen that shows how many consumers approached them and how much money they made that round. Consumers first decide whether they want to leave the market, or choose one of the three experts. Consumers do not observe an expert's ability level, but they can identify each expert, which allows for reputation building. After choosing an expert or leaving the market, consumers wait to be treated and proceed to a summary screen that shows which expert they approached in this round (as well as in all previous rounds), what prices the expert chose, and how much money they earned.

After 10 rounds, all subjects are informed that experts now have the opportunity to invest into their diagnostic precision. There are two treatments. In *Skill*, subjects either learn that experts can pay 10 Coins each round to increase their diagnostic precision to 90%. In *Algorithm*, experts can pay 10 Coins each round to rent an algorithmic decision aid that increases the expert's maximum diagnostic precision to 90% if used correctly. For the algorithmic decision aid, consumers also learn that experts are not forced to use the system, but can choose to ignore it. All subjects know that consumers pay for the investment by automatically paying 10 Coins more per treatment if they choose to approach an investing expert. Then, subjects complete another 15 rounds. Experts first decide whether they want to invest, then choose their price vector, and proceed to diagnosis and treatment. Consumers observe each expert's investment decision. Otherwise, nothing changes. Upon completing the credence goods experiment, subjects proceed to a short post-experimental questionnaire and answer a battery of demographic questions as well as a question about their risk attitudes (Dohmen et al., 2011).

Predictions

2.2 Phase 1 Equilibria

To derive equilibria, we differentiate between the three possible markup scenarios: (i) $\bar{p}-\bar{c} > \underline{p}-\underline{c}$; (ii) $\bar{p}-\bar{c} < \underline{p}-\underline{c}$; (iii) $\bar{p}-\bar{c} = \underline{p}-\underline{c}$. First note that in any equilibrium where the expert chooses $\mathbf{P}^s$ or $\mathbf{P}^m$ and consequently always implements only the high quality or low quality treatment, diagnostic precision does not matter and prices as well as profits are the same as in the model with transparent expert abilities. That is because consumers anticipate experts to follow their monetary self-interest. Under $\mathbf{P^e}$, experts signal honesty, because their markups for both treatments are the same. We first derive consumer profits under $\mathbf{P}^s$ and $\mathbf{P}^m$.

When (i) the profit margin for the HQT is larger than for the LQT, all expert types always recommend the HQT. Therefore, consumer problems are always solved. Profits are

$$\pi^m = v - \bar{p}. \tag{1}$$

When (ii) the profit margin for the LQT is larger than for the HQT, experts always recommend the low quality treatment, irrespective of ability. Consumers infer the markups from prices and anticipate undertreatment. Profits are

$$\pi^s = (1-h)v - \underline{p}. \tag{2}$$

Finally, there are equal markup scenarios. Here, heterogeneous expert abilities affect consumer payoffs, because lower diagnostic precision leads to more mistakes when treating consumers honestly. Under transparency, consumers anticipate different problems depending on expert type:

$$\pi^e = (1-h+hk)v - \bar{p} + (h+k-2hk)\Delta p \tag{3}$$

Which markup scenario do experts choose? With perfect diagnosis and full transparency, Dulleck and Kerschbamer (2006) show that under verifiability, experts always efficiently serve consumers by setting equal markups. With diagnostic uncertainty, that is no longer viable, because credible commitment to honesty does not imply that a consumer is sufficiently treated. Under competition, experts choose prices that maximize expected consumer income. Hence, the boundaries determining experts price strategy are:

$$\pi^m \geq \pi^e \, \text{when} \, h \geq \frac{k\Delta p}{(1-k)v - (1-2k)\Delta p} := h_i^m. \tag{4}$$

and

$$\pi^s \geq \pi^e \, \text{when} \, h \leq \frac{(1-k)\Delta c}{kv + (1-2k)\Delta p} := h_i^s. \tag{5}$$

Under full obfuscation, however, consumers do not know whether to condition their expected payoff on z or q. Therefore, they need to incorporate both possibilities, and profits change to

$$\pi^e = (1-h+hq(1-\gamma)+\gamma hz)v - \bar{p} + (h-2hq(1-\gamma)-\gamma 2hz+q(1-\gamma)+\gamma z)\Delta p. \tag{6}$$

The boundaries determining experts price strategy are:

$$\pi^m \leq \pi^e \text{ when } h \leq \frac{\Delta p(q-\gamma q+\gamma z)}{v(1-q(1-\gamma)-\gamma z)-\Delta p(1-2q(1-\gamma)-\gamma 2z)} := h_o^m. \tag{7}$$

and

$$\pi^s \leq \pi^e \text{ when } h \geq \frac{\Delta p(q-\gamma q+\gamma z-1)}{v(q(\gamma-1)-\gamma z)-\Delta p(1-2q(1-\gamma)-\gamma 2z)} := h_o^s. \tag{8}$$

Note that in comparison to the transparent scenario, expected consumer profits from approaching the high-ability expert decrease. This makes a straightforward argument why high-ability experts are incentivized to signal their true ability type.

We now turn towards three possible equilibria. We assume that consumers do not meaningfully learn about an expert's type over the 10 rounds of Phase 1. That is because (1) low-ability experts can always mimic high-ability experts, (2) consumers usually cannot verify dishonest overtreatment because they do not observe their problem, and (3) diagnostic uncertainty makes it even more difficult for consumers to discern between dishonesty and ability, e.g., in the case of undertreatment. Therefore, from the expert's perspective, their personal ability type is inconsequential. Given full obfuscation, both expert types charge the same prices and earn the same profits in equilibrium (see also Liu et al., 2020). If the high-ability expert would post a price vector that increases expected consumer profits, the low-ability expert could simply copy it. This works as long as obfuscation is strong enough such that consumers are sufficiently unlikely to believe that an expert following the high-ability expert's optimal pricing path is, in fact, a high-ability expert. In any equilibrium where experts choose $\mathbf{P}^s$ or $\mathbf{P}^m$ and consequently always implement only the high quality or low quality treatment, diagnostic precision does not matter.

In a first equilibrium, experts may always choose the HQT and charge $\mathbf{P}^m$, irrespective of type. This only manifests as long as consumer beliefs $\mu(\mathbf{P})$ that the expert is a high ability expert are low enough. Otherwise, the high-ability expert would be able to exploit their skill advantage by committing to treat consumers truthfully and offer higher profits. Hence, the expected loss from inadvertently approaching the low-ability expert who charges $\mathbf{P}^e$ (as compared to approaching an expert who charges $\mathbf{P}^m$) must be greater than or equal to the expected benefit from approaching the high-ability expert who charges $\mathbf{P}^e$.[2] Then, consumer beliefs that the expert is a

[2]This only works for values of h and z under which the transparent high-ability expert would be able to charge $\mathbf{P}^e$. Otherwise, signalling better diagnostic precision has no impact on the expert's

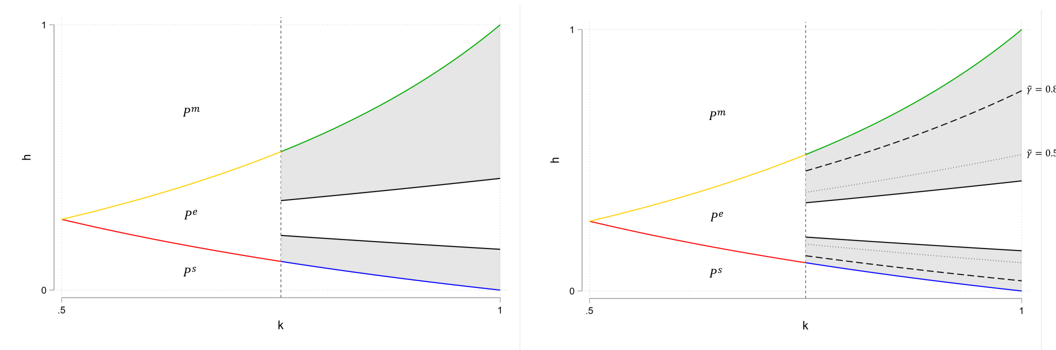

Figure 1. Left: Price setting under fully transparent and fully obfuscated expert abilities. The shaded areas represents the high-ability expert's loss. Due to consumer uncertainty, the high-ability expert cannot choose the profit-maximizing price vector $\mathbf{P}^e$ as often as they would want to. Variables for this figure: $q = 0.5$, $\gamma = 1/3$. Right: Variations for different consumer beliefs $\widetilde{\gamma}$.

high-ability expert do not exceed:

$$\widetilde{\gamma}^m := \frac{vh(1-q) - (h - 2hq + q)\Delta p}{(z-q)(vh - 2h\Delta p + \Delta p)}. \tag{9}$$

Equilibrium 1. Both expert types charge $\mathbf{P}^m$, always perform the major treatment, consumers believe $\mu(\mathbf{P}) \in [0, \widetilde{\gamma^m}]$ if the expert charges $\mathbf{P}^e$, meaning $\underline{p} - \underline{c} = \bar{p} - \bar{c}$, and consumers always visit the expert. Note that this also implies that the high-ability expert *could* attract more consumers and thus increase profits in a scenario with only partial obfuscation by increasing consumer beliefs such that $\widetilde{\gamma} > \widetilde{\gamma}^m$ for $h \in [h_o^m, h_H^m]$.

Equilibrium 2. The analogous equilibrium for the case where experts always choose the minor treatment implies:

$$\widetilde{\gamma}^s := \frac{-vhq - (h - 2hq + q - 1)\Delta p}{(z-q)(vh - 2h\Delta p + \Delta p)}. \tag{10}$$

Both expert types charge $\mathbf{P}^s$, always perform the minor treatment, consumers believe $\mu(\mathbf{P}) \in [0, \widetilde{\gamma^s}]$ if the expert charges $\mathbf{P}^e$, meaning $\underline{p} - \underline{c} = \bar{p} - \bar{c}$, and consumers always visit the expert. The high-ability expert could improve their profits in a scenario with only partial obfuscation by changing consumer beliefs such that $\widetilde{\gamma} > \widetilde{\gamma}^s$ for $h \in [h_o^s, h_H^s]$.

Equilibrium 3. Finally, there are equal mark-up equilibria. Both expert types

price setting.

charge $\mathbf{P}^e$ and consumers believe $\mu(\mathbf{P}) \in [0, \widetilde{\gamma}]$ because abilities are fully obfuscated and beliefs are fully determined by nature. Consumers always visit the expert. Contrary to the other two equilibria, treatment behavior is not obvious. We know that for $h \in [h_L^s, h_L^m]$, both expert types always follow their own diagnosis. However, within the gray area (see Figure 1) as well as at any point between h_L and h_H, the incentives of the two expert types to follow their diagnosis differ. That is because the high-ability expert may have options to deliver better treatments for the consumer, whereas the low-ability expert, knowing about their small diagnostic precision, could be incentivized to completely rely on one treatment and thereby minimize diagnostic mistakes. For the consumer, treatment choices do not make a difference, because under full obfuscation, they cannot infer anything from prices or experience. Without any additional constraints, treatment behavior could manifest in a variety of ways. Thus, we propose:

Proposition 1. *When expert abilities are fully obfuscated, the following scenarios arise:*

$$\left\{ \begin{array}{ll} \mathbf{P}^s & \text{if } h \in [0, h^s], \\ \mathbf{P}^e & \text{if } h \in [h^s,\, h^m], \\ \mathbf{P}^m & \text{if } h \in [h^m, 1] \end{array} \right\} \tag{11}$$

2.3 Phase 1 Hypotheses

We can now insert the experimental parameters to derive our hypothesis. In Phase 1, consumers believe that an expert is of high ability type with $\gamma = 1/3$. Expected consumer income equals $\pi^m = 50$ under $\mathbf{P}^m$, $\pi^s = 10$ under $\mathbf{P}^s$, and $\pi^e = 45.67$ under $\mathbf{P}^e$. This follows from consumers observing the mark-up incentives for experts and adjusting their expectations accordingly. Therefore, experts always charge $\mathbf{P}^m$ in equilibrium irrespective of type, with $h = 0.4 > h^m = 0.34$ and $\gamma = 1/3 < \widetilde{\gamma}^m = 0.59$. There is no incentive to deviate from implementing the high-quality treatment, meaning experts never under-treat clients, who therefore always enter the market. Experts over-treat clients with a probability of $1 - h = 0.6$. High-ability experts cannot differentiate themselves from low-ability experts, and price-setting therefore does not differ between types.

Hypothesis 1: Experts maximize expected consumer income by choosing $\mathbf{P}^m$ irrespective of type.

Hypothesis 2: There is no difference between high-ability expert and low-ability expert behavior.

Hypothesis 3: Consumers always enter the market.

2.4 Phase 2 Hypotheses: Skill

In *Skill*, experts can increase their diagnostic precision to $k = 0.9$ by paying $d = 10$ before choosing prices. Consumers observe whether an expert invested. Hence, by in investing, an expert's prior personal ability type becomes obsolete, and all investing experts are the same type. Expected client income from approaching an expert who invested into their diagnostic precision equals $\pi^e_{inv} = 57$. Comparing π^e_{inv} to π^m, π^s and π^e reveals a straightforward incentive for experts to invest. Under full transparency, the low ability experts offers consumers an expected income of 40 under $\mathbf{P^e}$, and the high-ability experts offers an expected income of 57. Hence, even if consumers infer something about the expert's type when they invest, no expert type can out-compete an investing expert. At most, consumers are indifferent between approaching the investing and the not-investing transparent high-ability expert, when $\widetilde{\gamma} = 1$. Otherwise, they prefer any expert who invested, regardless of their prior ability type. Assuming that high-ability experts cannot signal their ability type with certainty by foregoing investments, consumer beliefs about an expert's type are inconsequential, and we propose:

Hypothesis 4: Low-ability experts in *Skill* always invest to improve their diagnostic ability.

Hypothesis 5: Low-ability experts in *Skill* maximize expected consumer income by choosing $\mathbf{P}^e$.

Hypothesis 6: High-ability experts in *Skill* always invest to improve their diagnostic ability.

Hypothesis 7: High-ability experts in *Skill* maximize expected consumer income by choosing $\mathbf{P}^e$.

2.5 Phase 2 Hypotheses: Algorithm

In *Algorithm*, experts can rent an algorithmic decision aid that increases their diagnostic precision to $k = 0.9$ by paying $d = 10$ before choosing prices. Compared to

Skill, this setup exhibits some crucial differences. First, consumers know that experts can choose to forego the decision aid's information even after paying for the algorithm. Second, because consumers do not have insight into the expert's diagnostic process, they may assume some positive correlation between ability type and decision aid efficacy. Hence, consumers still consider an expert's personal precision when anticipating their expected income. We argue that this allows high-ability experts to differentiate themselves from low-ability experts, and escape the pooling equilibrium derived earlier.

Looking at expert incentives, first note that once uncertainty is reduced and consumer beliefs exceed certain thresholds $\widetilde{\gamma}$, pooling equilibria break down because high-ability experts can e.g. change their pricing strategy in return for their increased estimated diagnostic precision. The same is true for the low-ability expert, but reversed. Lower uncertainty decreases their potential profits. Thus, as long as there is uncertainty, high-ability experts experience a penalty, whereas low-ability experts can charge more. Figure 1 illustrates how different levels of $\widetilde{\gamma}$ change price setting. Without investments, high-ability experts have obvious incentives to persuade consumers of their diagnostic type.

With investments, we have shown above that a transparent high-ability expert would be indifferent between investing and not investing because in both cases, they offer consumers an expected income of $\pi^e = 57$ after choosing $\mathbf{P^e}$. Introducing the aforementioned uncertainty around decision aids changes that. Because consumers cannot be certain that experts utilize the algorithmic decision aid (i) at all or (ii) effectively, expected consumer income after investing changes to $\pi^e_{alg,inv} \leq 57$. Then, depending on the level of uncertainty as well as consumer beliefs about an expert's type, it may be profitable for high-ability experts to forego investments, if they can plausibly signal their type to consumers who judge the benefits of the decision aid as uncertain. Yet, if that is the case, low-ability experts may also have incentives to once again imitate high-ability experts, which is initially hidden to consumers because of obfuscated expert abilities. Note, however, that in order to imitate high-ability experts, low-ability experts now also rely on $\mathbf{P^e}$. In a one-shot game, this is irrelevant, because there is no consumer belief formation. In *Phase 1*, it is irrelevant because the pooling equilibrium implies $\mathbf{P^m}$, which masks the low-ability expert's abilities. In *Phase 2*, however, imitating low-ability experts cannot hide their relative diagnostic imprecision due to consumer expectations of being treated honestly (the high-ability expert's strategy). Therefore, we now have to ask whether it is profitable for low-ability experts to imitate high-ability experts despite Bayesian consumers who update their beliefs, or not. Following Mimra et al. (2016), we focus on behavior when consumers coordinate their behavior ex ante.

First, to allow high-ability experts to try and differentiate themselves from low-ability experts by foregoing the algorithm, and thereby to allow low-ability experts to imitate them by foregoing the algorithm, consumers need to believe that an expert who does not invest in round 11 is of high-ability type with $\mu(\mathbf{P}) \in [\widetilde{\gamma^m}, 1]$ where $\widetilde{\gamma^m} = 0.59$. Consumers then use Bayes's rule to update their beliefs $\mu(\mathbf{P})$ each round. There are three possible observations: (a) undertreatment (10 Coins), (b) overtreatment or efficient treatment of big problem (50 Coins), and (c) efficient treatment of small problem (90 Coins). An honest high-ability (low-ability) expert has respective observation probabilities of 0.1 (0.2), 0.45 (0.5) and 0.45 (0.3). Let n be the number of relevant a-signals, m be the number of relevant b-signals, and o be the number of relevant c-signals. Approaching a low-ability expert is an L-event, approaching a high-ability expert is a H-event. Given any sample of observations, consumers calculate the posterior probability:

$$Pr(L|n,m,o) = \frac{Pr(n,m,o|L)Pr(L)}{Pr(n,m,o|L)Pr(L) + Pr(n,m,o|H)Pr(H)}. \tag{12}$$

$Pr(L)$ is the prior probability for event L that is determined by consumer beliefs at the beginning of round 11 $\mu(\mathbf{P}) \in [\widetilde{\gamma^m}, 1]$. Using this, we can simulate a probability distribution of consumer beliefs. To demonstrate, let $\mu(\mathbf{P}) = \widetilde{\gamma} \in \{0.6, 0.8\}$. That is, consumers believe that their expert has high abilities with a probability of 60% or 80%. At this point, consumers would rather approach a not-investing expert with $\mathbf{P^e}$ under full obfuscation than a $\mathbf{P^m}$-expert who always successfully treats the problem because $\widetilde{\gamma} > \widetilde{\gamma^m}$. Figure 2 shows average consumer beliefs that a representative expert is of low-ability type over time, using 10.000 simulations. Here, we assume that low-ability experts fully imitate high-ability experts by choosing $\mathbf{P^e}$ and following their diagnosis.

It is clear that consumer beliefs change substantially over time, which constraints the imitation of high-ability experts.[3] A second possibility is that low-ability experts imitate high-ability experts in price setting and investment behavior, but always

[3]On the equilibrium path, consumers do not switch away from an expert in Phase 1 because there is no undertreatment, and expect some undertreatment in Phase 2 as long as it is optimal to approach a $\mathbf{P^e}$-expert. Therefore, consumers gather repeated observations from an individual expert.

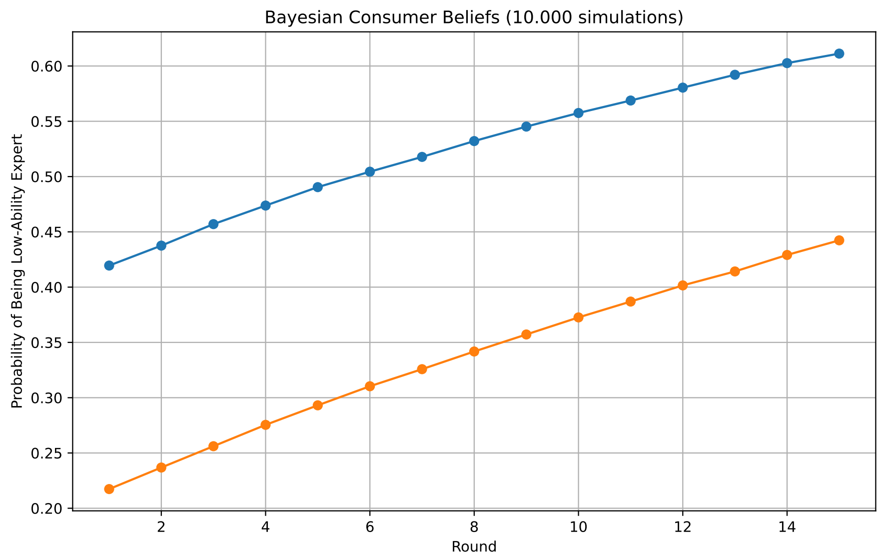

Figure 2. Consumer Belief Updating with $Pr(L) = 0.4$ and $Pr(L) = 0.2$.

choose the HQT in order to somewhat compensate their diagnostic disadvantage. As shown in Figure 6 in the appendix, this strategy very quickly reveals the expert's ability type. Having demonstrated the limits for low-ability experts to imitate high-ability experts, we now turn towards experts' optimal strategies and use an example to demonstrate that separating equilibria exist.

First, under which conditions does the high-ability expert deliberately forego the decision aid? Let $w \in [0, 1]$ be the estimated probability from the consumer that an investing expert does not (correctly) utilize the decision aid after renting it. Then, expected consumer profit equals $\pi^e_{alg,inv} = (1 - w) * \pi^e_{inv} + w * (\pi^e_o - 10) = (1 - w) * 57 + w * (30 + 17 * \frac{1}{3})$. In the event of w, consumers pay an additional d without gaining diagnostic improvements. Because experts invest, there is no signal, and $\widetilde{\gamma} = \gamma = 1/3$. On the other hand, if the expert does not invest, they influence consumer beliefs $\widetilde{\gamma} \in [\frac{1}{3}, 1]$. Expected consumer profits are $\pi^e_{alg,ninv} = 40 + 17 * \widetilde{\gamma}$. High-ability experts forego the decision aid along:

$$\pi^e_{alg,ninv} \geq \pi^e_{alg,inv} \text{ when } w \geq 0.797 * (1 - \widetilde{\gamma}) := w_{ninv}, \text{ or} \tag{13}$$

$$\pi^e_{alg,ninv} \geq \pi^e_{alg,inv} \text{ when } \widetilde{\gamma} \geq 1 - 1.255w := \widetilde{\gamma}_{ninv} \tag{14}$$

In a one-shot setting, low-ability experts imitate high-ability experts. With repeated

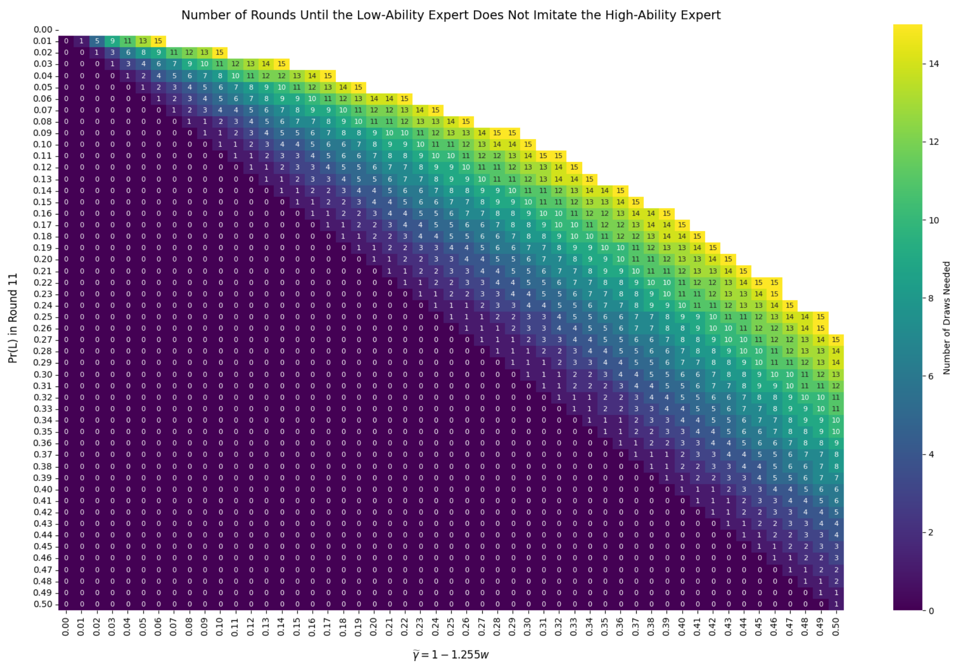

Figure 3. Heatmap of the number of rounds low-ability experts can imitate high-ability experts with a maximum of 15 rounds, given different values of $Pr(L)$ and $\widetilde{\gamma}_{ninv}$. For any point "0", high-ability experts cannot strategically forego the algorithm because consumers are not persuaded by the signal.

interactions, imitation becomes less profitable. To demonstrate the existence of separating equilibria, let $w = 0.1$. In that case, for an expert to deliberately forego the decision aid, consumers must believe that $Pr(L) < 0.125$, i.e. $\widetilde{\gamma} \in [0.875, 1]$. How long can the low-ability expert sustain their charade if they treat consumers honestly? For $Pr(L) \in \{0.1, 0.05, 0.01\}$, consumer beliefs $\widetilde{\gamma}$ fall below the threshold after two, eight and twenty-four rounds respectively (see Figure 3). The ability to imitate hinges on consumer priors and $\widetilde{\gamma}_{ninv}$. If $Pr(L)$ is extremely low, meaning consumers take non-investments as a very strong signal that an expert is of high-ability type, low-ability experts can imitate high-ability experts for a long time. In less extreme cases, that duration quickly falls, as exemplified by the 2 rounds under $Pr(L) = 0.1$. Now, assuming that $Pr(L) = 0.1$, what is the low-ability expert's income if they imitate the high-ability expert? Is it worth it?

Consumers ex ante coordinate on the low-ability expert. Because all experts choose $\mathbf{P^m}$ and the HQT in Phase 1, the expert does not lose any consumers and treats all three of them. In round 11, the expert can choose to forego the decision aid, and successfully imitate a high-ability expert. Then, each round the expert truthfully treats their consumers, their average profit amounts to $3 * 40 = 120$ Coins. After two rounds, consumers realize that the expert is not of high-ability type, and thus treated them sub-optimally. Therefore, they leave for either an investing expert, or another non-investing expert. Thus, the expert earns $2 * 120 + 13 * 15 = 435$ Coins over 15 rounds. If the expert had invested, they could have offered consumers an expected income (given $w = 0.1$) of $54, 98$ Coins. In order to leave, consumers must by assumption expect another expert to offer a better deal. Therefore, they do not leave for another investing expert. If another expert does not invest, consumers do not know whether that expert is the high-ability or the low-ability one. Because consumers do not switch towards an investing expert, both other experts have incentives to forego the investment. Therefore, expected consumer income when switching towards one of the two non-investing experts is: $\pi^e = 48.5$ Coins, which is lower than their expected income when staying. The low-ability expert keeps their consumers, and earns $15 * 120 = 1800$ Coins, which is more than the 435 Coins from imitating a high-ability expert.

High-ability experts, on the other hand, are always incentivized to forego the decision aid along w_{ninv} because consumer confidence in their ability type increases, on average, over time, sustaining the profitability of the strategy. Then, for any $w > 0$, non-investing high-ability experts make consumers a better offer than investing experts. If the high-ability expert invested, low-ability consumers could try to poach consumers for a short time by acting like a high-ability expert. This is not sustainable, but may still lead to income losses. Furthermore, if the high-ability expert invests, they offer the same deal as all other investing experts. Under the assumption that consumers do not switch away from an expert when they are indifferent between them and their alternatives, this does not affect expert behavior. However, foregoing the decision aid is robust to relaxing that assumption, and therefore beneficial to the high-ability expert. It is also robust to consumers randomly exploring at the beginning of *Phase 2*. Hence, separating equilibria exist. Note, however, that there are many instances where it is not profitable for high-ability experts to forego the decision aid (see Figure 3). In particular, low values of w as well as too low or too high values of $Pr(H)$, i.e. $\widetilde{\gamma}$, in round 11 prohibit high-ability signalling. If $\widetilde{\gamma} < \widetilde{\gamma}_{ninv}$, high-ability experts cannot send a strong enough signal, and therefore never forego the decision aid. If $\widetilde{\gamma}$ is too high, consumers' priors are so strong that low-ability experts sustain imitation.

We therefore hypothesize:

Hypothesis 8: Low-ability experts in *Algorithm* invest to rent the algorithmic decision aid.

Hypothesis 9: Low-ability experts in *Algorithm* maximize expected consumer income by choosing $\mathbf{P^e}$.

Hypothesis 10: Less high-ability experts than low-ability experts in *Algorithm* invest to rent the algorithmic decision aid.

Hypothesis 11: High-ability experts in *Algorithm* maximize expected consumer income by choosing $\mathbf{P^e}$.

3 Results

Table 1 provides summary statistics for the two periods. We first analyze subject behavior and market efficiency in the first 10 rounds without diagnostic investments. In contrast to the model's prediction, experts do not generally maximize expected consumer income by choosing $\mathbf{P}^m$. The most popular price vector $\mathbf{P}^s$ has the highest prices and lowest expected income for consumers (see also Figure 8 in the Appendix). This replicates a common effect from prior credence goods experiments *without* diagnostic uncertainty and obfuscated ability heterogeneity: under verifiability, experts rarely choose equal mark-up price vectors, and instead opt for the price vector that splits the gains of trade most equally if expert behave honestly (Dulleck et al., 2011; Kerschbamer et al., 2017). Consumers do not adequately punish undertreatment, and therefore fail to enforce price competition between experts. This leads to efficiency losses, as roughly 30% of big problems are treated with the LQT. Still, compared to other experiments with expert price setting, both $\mathbf{P}^m$ and $\mathbf{P}^e$ are relatively common. In fact, experts shift towards the predicted price vector $\mathbf{P}^m$ over time (logistic panel regression odds ratio: 1.08, p = 0.02, CI (95%): [1.01, 1.14]). Furthermore, as predicted by the model, high-ability experts ($\mathbf{P}^m/\mathbf{P}^s/\mathbf{P}^e = 0.29/0.39/0.31$) set the same prices as low-ability experts ($\mathbf{P}^m/\mathbf{P}^s/\mathbf{P}^e = 0.28/0.41/0.31$). High-ability experts cannot exploit their diagnostic advantage due to ability obfuscation. Overall, inefficiencies from over- and undertreatment are relatively common (see Figure 7), but do not deter consumers from entering the market. Only 5% of consumers choices fall on the outside option. This makes the market much more efficient than standard laboratory experiments with perfect diagnostic, suggesting a potential upside of

	Phase 1	Phase 2 Skill	Phase 2 Algorithm
Market Entry Rate	0.95	0.98	0.98
Efficiency (relative)	0.61	0.65 (0.75)	0.63 (0.74)
Undertreatment	0.29	0.23	0.25
Overtreatment	0.55	0.43	0.49
Efficient Treatment	0.52	0.64	0.60
Efficient Provision	0.40	0.59	0.53
Total Expert Surplus	132	133	133
Total Client Surplus	131	138	133
Price Vectors	$\mathbf{P}^m$ 0.28	$\mathbf{P}^m$ 0.33	$\mathbf{P}^m$ 0.26
	$\mathbf{P}^s$ 0.41	$\mathbf{P}^s$ 0.41	$\mathbf{P}^s$ 0.40
	$\mathbf{P}^e$ 0.31	$\mathbf{P}^e$ 0.26	$\mathbf{P}^e$ 0.34
N	300	150	150

Table 1. Summary statistics by phase and treatment. *Efficiency* is calculated as (average actual profit - outside option)/(maximum possible average profit - outside option). *Relative Efficiency* refers to the market's efficiency without correcting for the efficiency loss due to investment costs. *Undertreatment* is the share of treatment choices where the consumer has a big problem but receives the LQT. *Overtreatment* is the share of treatment choices where the consumer has a small problem but receives the HQT. *Efficient Treatment* equals 1 when an expert makes the efficient treatment decision. *Efficient Provision* equals 1 when the expert receives the correct diagnostic signal and consequently chooses the efficient treatment.

uncertainty for market outcomes (see e.g. Tracy et al. (2023)).

3.1 Phase 2 – Expert Investments

Round 11 introduces either *skill investments* or an *algorithmic decision aid* as an external technological shock. Each round, experts can increase their maximum diagnostic precision to 90% by paying 10 Coins, which also automatically increases their prices by 10 Coins. Figure 4 shows investment shares for the two treatments and expert types in Phase 2. As predicted, there are no differences between expert types for skill investments. Over all 15 rounds, experts invest 64% of the time. Investments start at around 80%, but deteriorate significantly (see Table 2).
Thus, while popular, investment shares lie substantially below the predicted (and

Figure 4. Expert investment shares for *Skill* and *Algorithm* over the 15 rounds of Phase 2.

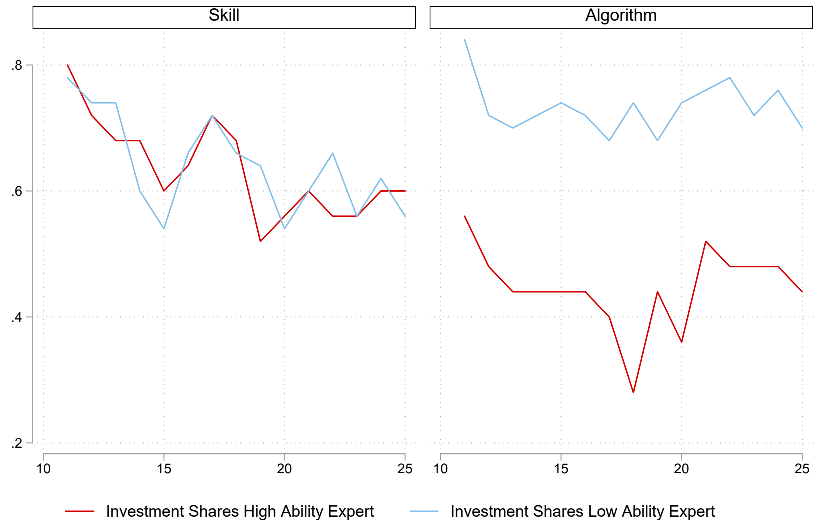

Table 2. Expert Investment Behavior in Phase 2

	Skill Investment	Skill Investment	Algorithm Investment	Algorithm Investment
High-Ability	-0.022	-0.035	-0.257***	-0.259***
	(0.081)	(0.079)	(0.067)	(0.071)
Round		-0.011**		-0.001
		(0.004)		(0.004)
Risk		0.024		-0.003
		(0.013)		(0.016)
Female		0.005		-0.016
		(0.083)		(0.061)
Age		-0.004		0.003
		(0.004)		(0.003)
N	1125	1125	1125	1125

This table reports marginal effects of panel logistic regressions using subject-level random effects and a cluster–robust VCE estimator at the matched group level (standard errors in parentheses). The dependent variable is a binary variable that equals 1 if the expert invests into the new diagnostic technology. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

optimal) 100%.

In *Algorithm*, investment shares clearly diverge between expert types. In round 11, 84% of low-ability and 56% of high-ability experts rent the algorithm ($\tilde{\chi}^2 = 6.91, p = 0.009$). Across all 15 rounds, 73% of low-ability expert choose to invest, whereas high-ability experts opt for the decision aid with a probability of only 45%. Logit panel regression with subject-level random effects and clustered standard errors at the matched group level confirm a large and significant negative effect of being a high-ability expert on investment shares (Table 2). This provides direct support for our main hypothesis. When an expert's initial ability type becomes inconsequential through skill investments, there are no differences between high- and low-ability experts. With technological decision aids, however, some high-ability experts try to signal their ability type by foregoing investments and offering lower prices. In line with that hypothesis, the algorithmic decision aid does not appear to be less desirable overall. In both treatments, aggregated investment shares are 64%. That is because, while high-ability experts exhibit a 19 percentage point drop in technology adoption, low-ability experts are qualitatively more likely choose the decision aid. Furthermore, investments into the algorithmic decision aid do not deteriorate over time. This suggests that subjects' do not predict consumers to be generally averse towards algorithms.

Expert Price Setting. Figure 5 shows the share of experts choosing a specific price vector depending on type, treatment and investment. We are primarily interested in potential strategic differences between experts who invest and those who forego investments. There are two potential effects. One, an expert may condition their pricing strategy on their current investment choice (within-subject variance). For instance, the standard model predicts a low-ability expert to change their pricing strategy from $\mathbf{P^m}$ to $\mathbf{P^e}$ after investing. Second, experts who invest a lot may differ from expert who do not invest or only invest occasionally (between-subject variance). This would, e.g., be consistent with some proportion of high-ability experts foregoing the investment choice to signal their type. Here, we do not expect any within-subject variance, because experts generally do not invest enough to generate the necessary variance in the data. We first focus on the second case. Table 3 shows panel regression results for price setting conditional on the number of rounds an expert chooses to invest. For low-ability experts, and *Skill* treatments, there are no differences. The number of investment choices does not predict price setting in either condition. For high-ability experts, there are large and significant effects of investing frequency on $\mathbf{P^s}$ and $\mathbf{P^e}$ in the *Algorithm* treatment. Those who often rent the algorithmic decision aid are more likely to choose $\mathbf{P^s}$ – the price menu that (1) has the highest consumer prices, (2) exhibits the highest expected value for experts and (3) splits the gains of trade relatively equally if experts behave honestly. On the other hand, high-ability experts who never or rarely invest into the decision aid are more likely to choose $\mathbf{P^e}$

Figure 5. Price setting depending on treatment, expert type and investment decision.

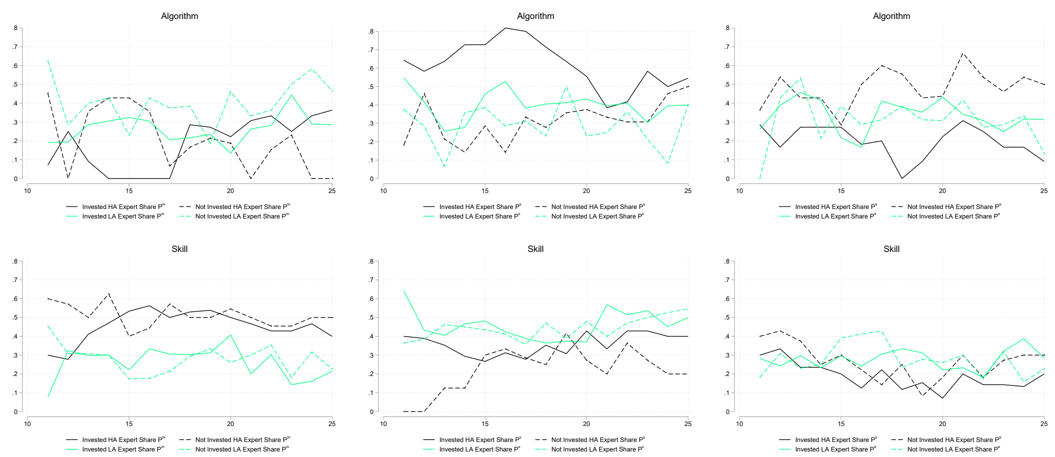

Table 3. Expert Price Setting Conditional on Investment Frequency

	Low-Ability Expert						High-Ability Expert					
	Skill			Algorithm			Skill			Algorithm		
	$\mathbf{P^m}$	$\mathbf{P^s}$	$\mathbf{P^e}$	$\mathbf{P^m}$	$\mathbf{P^s}$	$\mathbf{P^e}$	$\mathbf{P^m}$	$\mathbf{P^s}$	$\mathbf{P^e}$	$\mathbf{P^m}$	$\mathbf{P^s}$	$\mathbf{P^e}$
Investments	0.004	-0.002	-0.002	-0.009	0.001	0.013	0.001	0.006	-0.003	-0.004	0.045***	-0.044***
	(0.065)	(0.009)	(0.008)	(0.013)	(0.011)	(0.014)	(0.013)	(0.012)	(0.012)	(0.009)	(0.006)	(0.008)
N	750	750	750	750	750	750	375	375	375	375	375	375

This table reports marginal effects of panel logistic regressions using subject-level random effects and a cluster–robust VCE estimator at the matched group level (standard errors in parentheses). The dependent variable is a binary variable that equals 1 if the expert chooses the respective price vector. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

– the price menu that maximizes expected consumer income if they approach (a) a non-investing high-ability expert or (b) an investing expert.

These results are in line with our general proposition. In *Algorithm*, there appear to be systematic differences between high-ability experts who invest into new diagnostic technology and those who do not. Because high-ability experts can influence consumer perceptions, some of them strategically choose a low-investment path with equal markup prices. Those who invest a lot tend to opt for $\mathbf{P^s}$, which is again inconsistent with the standard model, but consistent with behavior in Phase 1 and prior empirical evidence about the prevalence of selfish expert preferences (Kerschbamer et al., 2017). In *Skill*, experts cannot differentiate themselves based on their original ability type. Therefore, all experts face the same incentives, and we would not expect two different "groups" that strategically choose either a low-investment or a high-investment strategy. Instead, experts should condition their behavior on their *current price menu*. To test this, we opt for a fixed effects logistic panel regression

Table 4. Expert Price Setting Conditional on Investment Decision

	Low-Ability Expert						High-Ability Expert					
	Skill			Algorithm			Skill			Algorithm		
	$\mathbf{P^m}$	$\mathbf{P^s}$	$\mathbf{P^e}$	$\mathbf{P^m}$	$\mathbf{P^s}$	$\mathbf{P^e}$	$\mathbf{P^m}$	$\mathbf{P^s}$	$\mathbf{P^e}$	$\mathbf{P^m}$	$\mathbf{P^s}$	$\mathbf{P^e}$
Invested	-0.134*	0.114	0.019	-0.173*	0.235***	-0.175***	-0.119	0.207**	-0.172*	0.009	-0.012	0.008
	(0.063)	(0.064)	(0.077)	(0.057)	(0.029)	(0.049)	(0.105)	(0.070)	(0.082)	(0.103)	(0.087)	(0.098)
N	510	555	540	540	645	630	210	210	255	195	255	210

This table reports marginal effects of fixed effects panel logistic regressions. The dependent variable is a binary variable that equals 1 if the expert chooses the respective price vector. Observations vary due to the omission of groups where investment shares predict prices perfectly. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

with a binary variable measuring whether an expert invested in a given round. This captures within-subject differences in price setting conditional on the expert's current investment choice. Results from a pooled regression suggest that generally, investing reduces the share of experts choosing $\mathbf{P^m}$ and $\mathbf{P^e}$ in favor of $\mathbf{P^s}$ (see Table 8). Separating regressions with regard to expert type and treatment reveals that these trends are present in all conditions, except high-ability experts in *Algorithm* (Table 4).
Thus, in our experiment, an investment into novel diagnostic technology that improves accuracy and simultaneously increases prices shifts experts towards the most expensive price menu with the lowest expected value for consumers. This tends to be true for both expert types across the two conditions, *except* for high-ability experts after the introduction of an algorithmic decision aid. In contrast, high-ability experts in *Algorithm* are the only group whose choices are explained by the absolute number of investments, with non- or low-investing experts opting for $\mathbf{P^e}$ instead of $\mathbf{P^s}$. Results therefore support the conjunction that high-ability experts diverge into two groups with the introduction of technological decision aids: one group that strategically under-invests, possibly to signal their ability type to consumers, and one group that follows the standard model's prediction by purchasing the algorithm. As predicted, this pattern only exists in *Algorithm*, but not in *Skill*.
Expert Undertreatment and Overtreatment. The prevalence of $\mathbf{P^s}$ may be an indication that experts build a reputation of honesty and can therefore offer to split the gains of trade equally. Table 5 shows that this is not generally the case. Consumers who approach a $\mathbf{P^s}$-expert are more likely to be undertreated, whereas $\mathbf{P^e}$ is – as theoretically predicted – associated with lower rates of undertreatment. This is true for both *Phase 1* and *Skill*. Only in *Algorithm*, the chosen price menu has no significant predictive power. Therefore, consumers cannot plausibly assume

Table 5. Undertreatment Conditional on Expert Price Setting

	Phase 1			Skill			Algorithm		
Undertreated	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$\mathbf{P^m}$	-0.084			-0.153**			-0.08		
	(0.054)			(0.054)			(0.065)		
$\mathbf{P^s}$		0.221***			0.252***			0.121	
		(0.041)			(0.043)			(0.066)	
$\mathbf{P^e}$			-0.152**			-0.131**			-0.048
			(0.049)			(0.047)			(0.074)
N	574	574	574	450	450	450	434	434	434

This table reports marginal effects of panel logistic regressions using subject-level random effects and a cluster–robust VCE estimator at the matched group level (standard errors in parentheses). The dependent variable is a binary variable that equals 1 if the consumers experiences undertreatment. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

experts to treat them honestly irrespective of prices. Investment choices do not predict undertreatment.

We now look at undertreatment from the expert's perspective. Each round, experts receive three diagnoses, and make three treatment choices. Because of diagnostic uncertainty, an expert may intent to undertreat a consumer, but does not actually do so because of a wrong diagnosis. Therefore, we define the intent-to-undertreat as experts who diagnose a big problem, but prescribe the LQT, irrespective of the consumer's actual underlying condition. The results (Table 6) confirm that across all treatments, experts cheating behavior is strongly influenced by their chosen price menu. Experts intent to undertreat consumers significantly more often under $\mathbf{P^s}$, and significantly less often under the two other menus. Investment choices are largely irrelevant. There are no differences between expert types.

For overtreatment, we document all relevant regression analyses in the Appendix. As expected, consumers experience more overtreatment when approaching an expert with $\mathbf{P^m}$ or $\mathbf{P^e}$ in Phase 1 (Table 10). For Phase 2, the results are similar, albeit less strong. In *Skill* but not *Algorithm*, consumers who approach an investing expert are also significantly less likely to be overtreated. Focusing on expert choices confirms the strong negative effect of $\mathbf{P^s}$ on overtreatment compared to both $\mathbf{P^m}$ and $\mathbf{P^e}$. Overall the results are in line with theory such that consumers should expect $\mathbf{P^s}$-experts to be much more likely to undertreat them. Across all rounds and conditions, intention-to-undertreat rates are 40% under $\mathbf{P^s}$, and only 12% and 11% for $\mathbf{P^m}$ and $\mathbf{P^e}$ respectively. On the other hand, intention-to-overtreat rates are only 22% under

Table 6. Expert Intent-To-Undertreat Conditional on Price Setting

	Phase 1			Skill			Algorithm		
Undertreated	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$\mathbf{P^m}$	-0.112***			-0.128**			-0.090*		
	(0.028)			(0.041)			(0.039)		
$\mathbf{P^s}$		0.191***			0.213***			0.199***	
		(0.031)			(0.056)			(0.066)	
$\mathbf{P^e}$			-0.095**			-0.097*			-0.149***
			(0.049)			(0.046)			(0.038)
Invested				-0.017	-0.026	-0.012	-0.039	-0.063*	-0.047
				(0.027)	(0.026)	(0.027)	(0.028)	(0.025)	(0.028)
N	1312	1312	1312	928	928	928	910	910	910

This table reports panel regressions using subject-level random effects and a cluster–robust VCE estimator at the matched group level (standard errors in parentheses). The dependent variable is the share of choices where an expert chooses a LQT for a diagnosed big problem. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

$\mathbf{P^s}$, and 48%/43% for $\mathbf{P^e}/\mathbf{P^m}$. Investment choices generally do not affect cheating intentions.

Consumer Choices. How do consumers react to undertreatment with diagnostic uncertainty? Table 7 shows that experiencing undertreatment strongly predicts consumer switching in the following round. This holds for both phases and treatments. Overtreatment appears to have no or only little effect on consumer choices. Furthermore, consumers do not appear to condition their switching decision on the expert's chosen price menu. Yet, looking at the number of consumers an expert attracts confirms that the former do exhibit strong preferences against $\mathbf{P^s}$, and for $\mathbf{P^m}$ (Table 11 in the Appendix). In line with consumer switching, undertreating consumers in the previous round also significantly reduces an expert's current number of consumers. Investments do not meaningfully affect an expert's ability to attract consumers. This may be one reason why we find consistent under-investment by experts. Finally, a three-way interaction of expert type, the expert's investment choice, and $\mathbf{P^e}$ suggests that high-ability experts in *Algorithm* attract significantly more consumers when they choose $\mathbf{P^e}$ and simultaneously forego the algorithmic decision aid. This result is in line with our theoretical prediction, and does not hold for either *Skill* or low-ability experts, nor for any other price menu. Thus, it provides supportive evidence for the hypothesis that high-ability experts may strategically alter their utilization of costly decision aids under obfuscation to influence consumer beliefs about their skill type.

Table 7. Consumer Switching Behavior

	Switched Away – Phase 1	Switched Away – Phase 1	Switched Away – Skill	Switched Away – Skill	Switched Away – Algorithm	Switched Away – Algorithm
Undertreated	0.377***	0.374***	0.256***	0.259***	0.208**	0.217**
	(0.051)	(0.049)	(0.067)	(0.059)	(0.068)	(0.063)
Overtreated	0.046	0.049	0.042	0.048	0.042	0.049*
	(0.026)	(0.026)	(0.038)	(0.041)	(0.028)	(0.024)
$\mathbf{P^m}$		0.024		-0.268		-0.229
		(0.092)		(0.253)		(0.195)
$\mathbf{P^s}$		-0.004		-0.156		-0.209
		(0.095)		(0.248)		(0.212)
$\mathbf{P^e}$		0.039		-0.364		-0.162
		(0.093)		(0.238)		(0.204)
Round		-0.008		-0.011*		-0.016***
		(0.006)		(0.005)		(0.004)
Risk		0.037***		0.027*		0.039***
		(0.009)		(0.012)		(0.011)
Invested				0.073*		-0.006
				(0.034)		(0.039)
Invested_LR				0.115*		0.021
				(0.046)		(0.036)
N	3600	3600	1050	1050	1050	1050

This table reports marginal effects of panel logistic regressions using subject-level random effects and a cluster–robust VCE estimator at the matched group level (standard errors in parentheses). The dependent variable is a binary variable that equals 1 if the consumer switched to a new expert in the current round. *Undertreated*, *Overtreated* and *Invested_LR* are lagged variables (one round). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4 Concluding Remarks

Technological breakthroughs have the potential to significantly alleviate inefficiencies on credence goods markets. While full automation of expert tasks is usually neither feasible, nor desirable or legal, algorithmic decision aids offer huge potential in complementing or sometimes even substituting human expertise. In this paper, we argue that two prevalent characteristics of real-life credence goods markets can inhibit the effective adoption of novel diagnostic technologies: diagnostic uncertainty and obfuscated heterogeneous expert abilities. Consumers know that experts cannot perfectly identify their problem, but also that some are better than others. Information asymmetries preclude consumers from accessing that information, hurting both their welfare and the welfare of high-ability experts. Technological shocks, like the invention of a new diagnostic medical test or large language models, offer an opportunity for experts to influence consumer perceptions of their abilities. That is because, one, consumers know that low-ability experts derive relatively larger benefits from decision aids, two, consumers tend to associate the utilization of decision aids with lower expert skills, and three, decision aids carry uncertainty. Furthermore, we show that in many situations, low-ability experts cannot afford to imitate the high-ability expert's decision to forego an algorithmic decision aid. Then, by strategically ignoring the decision aid, a high-ability expert can signal consumers their ability type, and simultaneously offer a more favorable price. In reality, the latter may represent (1) lower diagnostic costs for experts or (2) less services for the consumer. In the medical domain, for instance, it is common for patients to pay for additional or advanced tests that are not covered by insurances but increase the chance of a correct diagnosis.

We test this idea using an online credence goods experiment in which experts experience a technological shock after 10 rounds and thereby receive the option to costly increase their diagnostic precision. The experiment differentiates between skill investments, and algorithmic decision aids. Skill investments over-ride an expert's prior ability type, representing for example advanced or specialized training. Algorithmic decision aids complement the expert's diagnosis, without affecting their personal ability type. Therefore, experts who do not use or disregard the system may still rely on their own diagnostic skill. Our results show no differences between high-ability and low-ability expert for skill investments. With the algorithmic decision aid, however, high-ability experts are far less likely to pay for increased diagnostic precision. Auxiliary results suggest that investing and non-investing high-ability experts exhibit different pricing strategies, with those largely foregoing the decision aid predominantly relying on price menus that signal honesty. On the other hand, investing experts opt for prices that are more favourable to them. These results offer preliminary evidence that high-ability experts may strategically (under-)utilize technology to differentiate themselves from low-ability experts. They also suggest that the adoption of algorithmic decision aids could be conditional on the distribution of skills within a profession, and that negative consumer attitudes towards professionals with decision aids may be rooted in actual ability differences.

Overall, experts in our experiment substantially under-invest into new diagnostic technologies. This is one reason why efficiency gains are small to negligible. Investments do not meaningfully affect expert fraud, which may be partially driven by very high consumer participation rates. Throughout the experiment, 95%-98% of consumers enter the market, which speaks to the potential efficiency of uncertain credence goods markets under verifiability. Consumers react negatively towards undertreatment, but not overtreatment, and generally do not reward expert investments. In contrast to the predictions of the standard model, consumers also do not appear to reward expert prices that signal honesty. Instead, they avoid high prices with expert-incentives to under-treat, but approach lower prices with expert-incentives to over-treat. This strongly reduces investment incentives.

Our findings provide first evidence on the efficacy of algorithmic decision aids on credence goods markets with obfuscated, heterogeneous expert abilities. We hope they provide some useful orientation for future research. In particular, there is much to learn about the relationship between expert investments, technology, and consumer beliefs. For example, in this paper, we did not collect data on consumer beliefs. However, carefully isolating and quantifying the mechanism behind (a) consumer reactions towards experts with decision aids, and (b) high-ability experts shunning of decision aids, is crucial to better understand the transformative impact of technology on credence goods markets. This necessitates a better understanding of the beliefs and expectations of experts and consumers alike. One positive implication of our research is that high-ability experts may be dis-incentivized to offer additional tests with low marginal benefits. On the other hand, we find clear signs of under-investment, with no indication that consumers incentivize experts towards the optimal diagnostic path.

References

Alarie, Benjamin, Anthony Niblett, and Albert H Yoon, "How artificial intelligence will affect the practice of law," *University of Toronto Law Journal*, 2018, *68* (supplement 1), 106–124.

Aobdia, Daniel, Saad Siddiqui, and Andres Vinelli, "Heterogeneity in expertise in a credence goods setting: Evidence from audit partners," *Review of Accounting Studies*, 2021, *26*, 693–729.

Arkes, Hal R, Victoria A Shaffer, and Mitchell A Medow, "Patients derogate physicians who use a computer-assisted diagnostic aid," *Medical Decision Making*, 2007, *27* (2), 189–202.

Balafoutas, Loukas, Helena Fornwagner, Rudolf Kerschbamer, and Matthias Sutter, "Diagnostic uncertainty and insurance coverage in credence goods markets," 2020.

Bester, Helmut and Matthias Dahm, "Credence goods, costly diagnosis and subjective evaluation," *The Economic Journal*, 2018, *128* (611), 1367–1394.

Brenner, Lukas and Tobias Meyll, "Robo-advisors: a substitute for human financial advice?," *Journal of Behavioral and Experimental Finance*, 2020, *25*, 100275.

Brody, Howard, "Medicine's ethical responsibility for health care reform—the top five list," *New England Journal of Medicine*, 2010, *362* (4), 283–285.

Chan, David C, Matthew Gentzkow, and Chuan Yu, "Selection with variation in diagnostic skill: Evidence from radiologists," *The Quarterly Journal of Economics*, 2022, *137* (2), 729–783.

Chen, Daniel L, Martin Schonger, and Chris Wickens, "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.

Chen, Yongmin, Jianpei Li, and Jin Zhang, "Efficient liability in expert markets," *International Economic Review*, 2022, *63* (4), 1717–1744.

Cruickshank, PJ, "Patient rating of doctors using computers," *Social Science & Medicine*, 1985, *21* (6), 615–622.

Currie, Janet and W Bentley MacLeod, "Diagnosing expertise: Human capital, decision making, and performance among physicians," *Journal of labor economics*, 2017, *35* (1), 1–43.

Dai, Tinglong and Shubhranshu Singh, "Conspicuous by its absence: Diagnostic expert testing under uncertainty," *Marketing Science*, 2020, *39* (3), 540–563.

Daniels, Marcia and Steven A Schroeder, "Variation among physicians in use of laboratory tests II. Relation to clinical productivity and outcomes of care," *Medical care*, 1977, pp. 482–487.

Dawes, Robyn M, David Faust, and Paul E Meehl, "Clinical versus actuarial judgment," *Science*, 1989, *243* (4899), 1668–1674.

Daysal, N Meltem, Sendhil Mullainathan, Ziad Obermeyer, Suproteem K Sarkar, and Mircea Trandafir, "An Economic Approach to Machine Learning in Health Policy," *Univ. of Copenhagen Dept. of Economics Discussion, CEBI Working Paper*, 2022, (24).

Dietvorst, Berkeley J, Joseph P Simmons, and Cade Massey, "Algorithm aversion: people erroneously avoid algorithms after seeing them err.," *Journal of Experimental Psychology: General*, 2015, *144* (1), 114.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G Wagner, "Individual risk attitudes: Measurement, determinants, and behavioral consequences," *Journal of the european economic association*, 2011, *9* (3), 522–550.

Doi, Dimitria, Romulo Ribeiro do Vale, Jean Michell Correia Monteiro, Glauco Cabral Marinho Plens, Mario Ferreira Junior, Luiz Augusto Marcondes Fonseca, Sandro Félix Perazzio, Bruno Adler Maccagnan Pinheiro Besen, Arnaldo Lichtenstein, Leandro Utino Taniguchi et al., "Perception of usefulness of laboratory tests ordering by internal medicine residents in ambulatory setting: A single-center prospective cohort study," *Plos one*, 2021, *16* (5), e0250769.

Dulleck, Uwe and Rudolf Kerschbamer, "On doctors, mechanics, and computer specialists: The economics of credence goods," *Journal of Economic literature*, 2006, *44* (1), 5–42.

_ and _ , "Experts vs. discounters: Consumer free-riding and experts withholding advice in markets for credence goods," *International Journal of Industrial Organization*, 2009, *27* (1), 15–23.

_ , _ , and Matthias Sutter, "The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition," *American Economic Review*, 2011, *101* (2), 526–555.

Eastwood, Joseph, Brent Snook, and Kirk Luther, "What people want from their professionals: Attitudes toward decision-making strategies," *Journal of Behavioral Decision Making*, 2012, *25* (5), 458–468.

Frankel, Alexander and Michael Schwarz, "Experts and their records," *Economic Inquiry*, 2014, *52* (1), 56–71.

Groopman, Jerome E and Michael Prichard, *How doctors think*, Vol. 82, Houghton Mifflin Boston, 2007.

Hall, Stephen F, Colleen Webber, Patti A Groome, Christopher M Booth, Paul Nguyen, and Yvonne DeWit, "Do doctors who order more routine medical tests diagnose more cancers? A population-based study from Ontario Canada," *Cancer medicine*, 2019, *8* (2), 850–859.

Hilger, Nathaniel G, "Why don't people trust experts?," *The Journal of Law and Economics*, 2016, *59* (2), 293–311.

Jr, Joseph J Doyle, Steven M Ewer, and Todd H Wagner, "Returns to physician human capital: Evidence from patients randomized to physician teams," *Journal of health economics*, 2010, *29* (6), 866–882.

Kerschbamer, Rudolf, Matthias Sutter, and Uwe Dulleck, "How social preferences shape incentives in (experimental) markets for credence goods," *The Economic Journal*, 2017, *127* (600), 393–416.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, "Human decisions and machine predictions," *The quarterly journal of economics*, 2018, *133* (1), 237–293.

Litman, Leib, Jonathan Robinson, and Tzvi Abberbock, "TurkPrime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences," *Behavior research methods*, 2017, *49* (2), 433–442.

Liu, Fang, Alexander Rasch, Christian Waibel, and Marco A Schwarz, "The role of diagnostic ability in markets for expert services," 2020.

Mimra, Wanda, Alexander Rasch, and Christian Waibel, "Price competition and reputation in credence goods markets: Experimental evidence," *Games and Economic Behavior*, 2016, *100*, 337–352.

Miyakis, Spiros, Georgios Karamanof, Michalis Liontos, and Theodore D Mountokalakis, "Factors contributing to inappropriate ordering of tests in an academic medical department and the effect of an educational feedback strategy," *Postgraduate medical journal*, 2006, *82* (974), 823–829.

Mullainathan, Sendhil and Ziad Obermeyer, "Diagnosing physician error: A machine learning approach to low-value health care," *The Quarterly Journal of Economics*, 2022, *137* (2), 679–727.

Pesendorfer, Wolfgang and Asher Wolinsky, "Second opinions and price competition: Inefficiency in the market for expert advice," *The Review of Economic Studies*, 2003, *70* (2), 417–437.

Promberger, Marianne and Jonathan Baron, "Do patients trust computers?," *Journal of Behavioral Decision Making*, 2006, *19* (5), 455–468.

Rasch, Alexander and Christian Waibel, "What drives fraud in a credence goods market?–Evidence from a field study," *Oxford Bulletin of Economics and Statistics*, 2018, *80* (3), 605–624.

Sandefur, Rebecca L, "Legal advice from nonlawyers: consumer demand, provider quality, and public harms," *Stan. JCR & CL*, 2020, *16*, 283.

Schneider, Henry S, "Agency problems and reputation in expert services: Evidence from auto repair," *The Journal of Industrial Economics*, 2012, *60* (3), 406–433.

Schneider, Tim and Kilian Bizer, "Effects of qualification in expert markets with price competition and endogenous verifiability," *CEGE Discussion Papers, Number*, 2017.

_ **and** _ , "Expert qualification in markets for expert services: A Sisyphean Task?," *CEGE-Discussion Papers, Number*, 2017.

Schroeder, Steven A, Alan Schliftman, and Thomas E Piemme, "Variation among physicians in use of laboratory tests: relation to quality of care," *Medical Care*, 1974, pp. 709–713.

Shaffer, Victoria A, C Adam Probst, Edgar C Merkle, Hal R Arkes, and Mitchell A Medow, "Why do patients derogate physicians who use a computer-based diagnostic support system?," *Medical Decision Making*, 2013, *33* (1), 108–118.

Topol, Eric, *Deep medicine: how artificial intelligence can make healthcare human again*, Hachette UK, 2019.

Tracy, Dustin, Eric Schniter, and Vojtěch Zíka, "Uncertainty and Reputation Effects in Credence Goods Markets," *Available at SSRN 4513820*, 2023.

Wolf, James R, "Do IT students prefer doctors who use IT?," *Computers in Human Behavior*, 2014, *35*, 287–294.

Xue, Hao, Yaojiang Shi, Lei Huang, Hongmei Yi, Huan Zhou, Chengchao Zhou, Sarah Kotb, Joseph D Tucker, and Sean Y Sylvia, "Diagnostic ability and inappropriate antibiotic prescriptions: a quasi-experimental study of primary care providers in rural China," *Journal of Antimicrobial Chemotherapy*, 2019, *74* (1), 256–263.

Yeh, D Dante, "A clinician's perspective on laboratory utilization management," *Clinica chimica acta*, 2014, *427*, 145–150.

Zhi, Ming, Eric L Ding, Jesse Theisen-Toupal, Julia Whelan, and Ramy Arnaout, "The landscape of inappropriate laboratory testing: a 15-year meta-analysis," *PloS one*, 2013, *8* (11), e78962.

5 Appendix

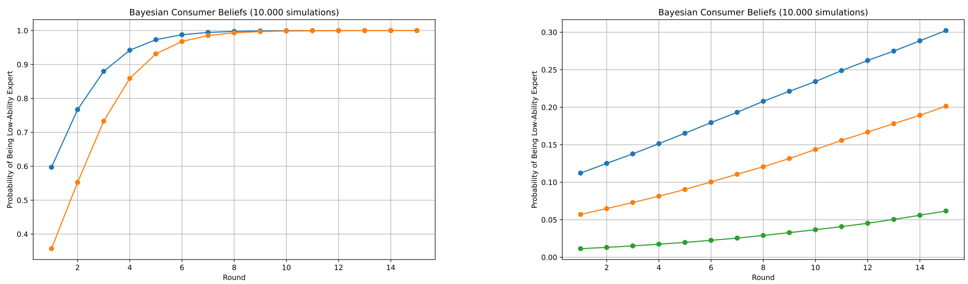

Figure 6. Left: Consumer Belief Updating with $Pr(L) = 0.4$ (blue) and $Pr(L) = 0.2$ (orange) when Low-Ability Experts only choose the HQT. Right: $Pr(L) \in \{0.1, 0.05, 0.01\}$.

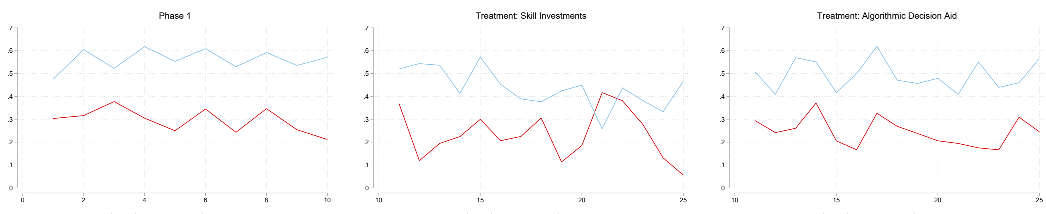

Figure 7. Share of undertreatment and overtreatment in (left) the first 10 rounds (middle) Phase 2 of *Skill* and (right) Phase 2 of *Algorithm*

.

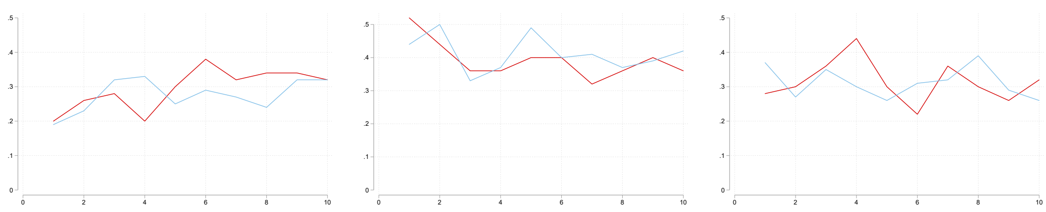

Figure 8. Expert price setting $\mathbf{P^m}$, $\mathbf{P^s}$ and $\mathbf{P^e}$ in Phase 1.

Table 8. Consumer Approaching Behavior

	$\mathbf{P^m}$	$\mathbf{P^s}$	$\mathbf{P^e}$
Invested	-0.122**	0.159***	-0.088*
	(0.039)	(0.031)	(0.040)
N	1,455	1,665	1,635

This table reports marginal effects of a pooled fixed effects panel logistic regression. The dependent variable is a binary variable that equals 1 if the expert chooses the respective price vector. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9. Frequencies of expert prices in Phase 2.

	Expert			
	HA Invested	HA Not Invested	LA Invested	LA Not Invested
Algorithm	$\mathbf{P^m}$: 18.6% $\mathbf{P^s}$: 61.1% $\mathbf{P^e}$: 20.4%	$\mathbf{P^m}$: 20.2% $\mathbf{P^s}$: 31.2% $\mathbf{P^e}$: 48.6%	$\mathbf{P^m}$: 26.3% $\mathbf{P^s}$: 40.3% $\mathbf{P^e}$: 33.3%	$\mathbf{P^m}$: 39.5% $\mathbf{P^s}$: 29% $\mathbf{P^e}$: 31.5%
Skill	$\mathbf{P^m}$: 44.9% $\mathbf{P^s}$: 35.7% $\mathbf{P^e}$: 19.3%	$\mathbf{P^m}$: 50.4% $\mathbf{P^s}$: 24.1% $\mathbf{P^e}$: 25.5%	$\mathbf{P^m}$: 25.9% $\mathbf{P^s}$: 46.2% $\mathbf{P^e}$: 27.9%	$\mathbf{P^m}$: 27.1% $\mathbf{P^s}$: 44.9% $\mathbf{P^e}$: 27.9%

Table 10. Overtreatment Conditional on Expert Price Setting

	Phase 1			Skill			Algorithm		
Overtreated	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$\mathbf{P^m}$	0.094*			0.120			0.111		
	(0.047)			(0.054)			(0.084)		
$\mathbf{P^s}$		-0.228***			-0.232**			-0.199**	
		(0.041)			(0.054)			(0.066)	
$\mathbf{P^e}$			0.119*			0.081			0.055
			(0.046)			(0.084)			(0.082)
N	856	856	856	652	652	652	664	664	664

This table reports marginal effects of panel logistic regressions using subject-level random effects and a cluster–robust VCE estimator at the matched group level (standard errors in parentheses). The dependent variable is a binary variable that equals 1 if the consumers experiences overtreatment. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11. Consumer Approaching Behavior

	Phase 1	Skill	Algorithm	Algorithm
Undertreated	-0.519**	-0.539***	-0.465*	-0.473*
	(0.169)	(0.212)	(0.220)	(0.221)
Overtreated	0.089	0.047	0.075	0.088
	(0.098)	(0.097)	(0.038)	(0.095)
$\mathbf{P^e}$	Baseline	Baseline	Baseline	Baseline
$\mathbf{P^m}$	0.574**	0.620**	1.174***	1.297**
	(0.174)	(0.225)	(0.283)	(0.402)
$\mathbf{P^s}$	-0.770***	-0.803***	-0.748***	-0.661**
	(0.167)	(0.215)	(0.138)	(0.229)
Not_Invested		0.059	-0.262	-0.386
		(0.276)	(0.166)	(0.209)
High-Ability				0.905
				(0.503)
Not_Invested × $\mathbf{P^e}$				0.393
				(0.375)
High-Ability × $\mathbf{P^e}$				-0.683
				(0.363)
Not_Invested × High-Ability × $\mathbf{P^e}$				0.849*
				(0.431)
N	1,350	1,125	1,125	1,125

This table reports results of a panel ordered logistic regressions using subject-level random effects and a cluster–robust VCE estimator at the matched group level (standard errors in parentheses). The dependent variable is an ordinal variable that captures the number of consumers (0 - 3) who approached the expert in the current round. *Undertreated* and *Overtreated* are lagged variables (one round). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$