Anonymous or personal? A simple model of repeated personalized advice

Christoph Schottmüller*

University of Copenhagen, Tilec

Preliminary version. Comments are very welcome.

June 1, 2018

Abstract

A consumer asks an expert repeatedly for advice. The expert's incentives are not aligend with the consumer's preferences because he can get a bonus if the consumer takes certain actions. Over time, the expert gets to know the consumer and is therefore able to give better advice (if he wants to do so). In simple equilibria, both – consumer and expert – benefit from the expert's learning if "learning" is such that the expert's best guess to what is the best advice for the consumer becomes more precise (in contrast to ruling out advice that is unlikely to fit the consumer with greater certainty). This gives a natural explanation for why consumers have a preference for personal advice and also for why most internet users do not use anonymization tools.

JEL codes:

Keywords: advice, cheap talk, privacy

1. Introduction

In many situations consumers ask better informed experts to guide their choices. This happens even in situations in which experts might have preferences over consumer choices that do not coincide with the consumers' preferences and it happens even in situations where it is difficult for the consumer to express his exact preferences precisely. For example, a consumer might ask employees of his bank for financial advice. The bank employee will usually earn a bonus if the consumer buys a certain investment product and often different investment products result in different boni for the adviser. There is no reason to believe that the product giving the highest bonus is also the product that

*email: christoph.schottmueller@econ.ku.dk, phone: +45 35 32 30 87

is most suitable for the consumer. Similar situations emerge in other retailing settings like consumer electronics or even cars.

Another example is internet search. A consumer submits a search term and relies on the answer of the search engine. Since some links are sponsored, the search engine has an incentive to display the sponsored links more prominently than links that better match the consumer's demands but are not sponsored. A third example, would be a minister (or manager) asking a civil servant (subordinate) to draft a certain legislative act or decree. Even if the civil servant has no political preferences of his own, he might be aware that a similar draft was written under a previous government and that handing this old draft to the minister would save him a considerable amount of time and effort. Again it is unlikely that this old draft exactly does what the minister wanted to achieve. As a final example, think of a physician patient relationship. The patient describes his symptoms and the phyisician prescribes a medication. Given the amount of lobbying by pharmaceutical companies, it is quite possible that the physician has a preference for a certain drug company or pharmaceutical product.

What are the common elements of these examples? A consumer asks an expert to guide his choice although the consumer cannot be sure of the expert's preferences. In none of the examples, there is a direct payment from the consumer to the expert which implies that the consumer has little possibilities to give the expert the right incentives. Furthermore, the consumer's communication of his preferences is complicated (due to the complicated nature of the subject and the consumer's ignorance that makes him ask for advice in the first place) and the task of the expert is difficult. Put differently, even if the expert tried to serve the consumer as well as he can, misunderstandings and errors would happen with some probability. In a static one shot game, we should therefore not expect any useful advice in any of these situations: By the one shot nature, an expert would optimally recommend the alternative giving him the (highest) bonus because the consumer has no possibility to punish this behavior. Knowing this, the consumer would then not even ask for advice as the recommendation would not bare any resemblance to his wishes. However, the examples above usually do not resemble a one shot game. Consumers repeatedly consult with the same financial adviser, use the same search engine, work with the same subordinates or visit the same physician. The repeated interaction – one could call it "building up a relationship" – has two interesting consequences: First, it is well known in game theory that cooperative behavior can be sustained in repeated interactions even if this behavior cannot be sustained in a static one shot game. Hence, meaningful advice could be possible because of the repeated nature of the advice situation. Second, the adviser could learn to interpret the wishes of the consumer. That is the ability of the adviser to give fitting recommendations is likely to improve over time. The reason is that both adviser and expert can observe how past recommendations played out, e.g. whether the consumer was happy with the product purchased (or tried to return it), whether the consumer clicked on the recommended link (and stayed on the website or bought something there afterwards), whether the draft was pushed forward or burried or whether the patient was cured. The success or failure of the recommendation gives the possibility to learn how to interpret future messages sent by the consumer.

It should be noted that the learning I have in mind is relationship specific. In particular, prior learning would be of little use to the consumer if he decided to switch experts. Although the consumer might to some degree also learn how to express his wishes, the biggest part of the learning appears to be on the side of the expert. This paper focuses therefore on a setting where only the expert learns and tries to answer several questions. The most basic one is whether an equilibrium with meaningful advice is possible. The answer is – unsurprisingly – in the affirmative. The expert will give partially useful advice in equilibrium because the consumer threatens to end the relationship (and therefore the expert's opportunity to collect boni) if he receives bad advice for a number periods. The main question is whether the consumer will benefit from the expert's learning. This is unclear as the consumer's outside option is not affected by the expert's learning, i.e. the expert could counteract his improved ability to give the right recommendation by recommending the product for which he receives a bonus more often. It will be shown that – under certain conditions – the consumer benefits nevertheless in a certain class of simple equilibria. The reason is a *value effect*. The more the expert learns about the consumer, the more valuable is the consumer to the expert in the sense that the expected discounted bonus stream from this consumer is higher. The expert will therefore loose more if the consumer ends the relationship and is therefore more inclined to give better advice in order to avoid exactly this. This leads to a testable prediction: the probability that a relationship ends now given that it has not ended yet is therefore lower the longer the relationship lasts.[1]

The sufficient condition for having the consumer benefit from the expert's learning is interesting in itself: It is satisfied if learning implies that the option that the expert identifies as most likely to fit the consumer's needs becomes relatively more likely. It is not satisfied if learning mainly implies that the least likely options become even less likely. The expert's learning can be detrimental for the consumer in the latter case.

The result that consumers benefit from the learning of the expert gives a natural explanation for a puzzle that emerged in the literature on privacy. People do not take even simple measures to anonymize their online activities. For example, most users use a search engine like Google directly instead of using an anonymized service that reroutes

[1]More precisely, the probability that the relationship ends this period is lower than it was m periods ago where $m \in \mathbb{N}$ is a number defined by the consumers equilibrium strategy.

their search requests through another server before submitting it to Google (thereby anonymizing it).[2] Privacy advocates emphasize that the more information the search engine has on a user, the higher the potential for exploitation (a simple exploitation method would be to display more sponsored links). The model shows that this is not the only effect. Due to the value effect, consumers also benefit from the search engine's learning. Staying anonymous would lead to lower consumer surplus in the model of this paper. This also explains why consumers might prefer to get advice from the same person, e.g. have the same financial advisor in their bank whenever they go there or stick with the same physician instead of switching every time they fall ill.

The model is related to the cheap talk literature started by Crawford and Sobel (1982) and surveyed in xxx. Therere are two notable differences. First, most of this literature is either static or deals with reputation concerns (Sobel, 1985; Benabou and Laroque, 1992; Park, 2005). Reputation issues are not treated in the setup of this paper. Second and more importantly, the cheap talk literature deals with a different misalignment of preferences. Typically, there is a one dimensional decision and the expert is biased in one direction, e.g. the expert prefers a somewhat higher decisions than the decision maker. The structure here is different as the expert simply has a preferred option that is independent from the consumer's optimal option. One implication of this structure is that no meaningful advice is possible in a static setting while this is obviously not true in the cheap talk liteature.

Literature: privacy (?), dynamic learning (?), more on dynamic cheap talk

2. Model

The model is an infinitely repeated game. In each period there are two options of which the *consumer (C)* has to take one. One of the two options fits C's needs and gives him therefore a payoff 1 while the other option gives him a payoff of 0. C's prior is that both options are equally likely to give him a payoff of 1.

The *expert (E)* receives a private and noisy signal concerning which option fits C's needs. More precisely, E's signal leads to a posterior where one option has probability $p_k > 1/2$ to fit C's needs and the other option has probability $1 - p_k < 1/2$ to fit C's needs. Without loss of generality I will call the option which is more likely to fit C's needs option 1. The precision of E's signal, p_k, is an element of a finite set $P = \{p_1, p_2, \ldots, p_n\}$ where $1/2 < p_1 < p_2 < \cdots < p_n < 1$. As E learns about C' needs over time, precision improves in the following way: Whenever E recommends the option fitting C's needs, precision improves from p_k to p_{k+1} (unless $p_k = p_n$ in which

[2] Many easy to use services like this exist, e.g. https://www.startpage.com or https://www.privatesearch.io.

case precision remains unchanged).[3]

The *expert's payoffs* are as follows: In every period, E has a *bonus option.* That is, E receives a bonus of 1 if he recommends this option to C while he receives a payoff of 0 otherwise. Each option is ex ante equally likely to be the bonus option and the identity of the bonus option is private information of E.

The *timing* is as follows. In each period, E privately observes his signal and the identity of his bonus option. Then E recommends an option to C. C follows this recommendation and period payoffs realize. In particular, C observes whether the recommendation fitted his needs or not. Afterwards C decides whether to *end* the game or to *continue.* If C ends the game, C receives an outside option V_O in the following period while E receives a zero payoff and no payoffs occur in later periods. If C continues, another period of the same game begins. Both players discount future payoffs with discount factor $\delta \in (0,1)$. Needs and bonus option are assumed to be independent from another and across periods.

To make the problem interesting, C's outside option should neither be too attractive nor too unattractive. For example, V_O should be lower than the value the consumer would get if he had a signal of precision p_n. If this was not satisfied, C would have the dominant strategy to end the relationship immediately. The outside option should also not to be too low. More precisely, I assume that V_O is higher than the value C gets if E recommends his bonus option in every period. If this did not hold, there would be a unique perfect Bayesian equilibrium in which C always continues and E always recommends his bonus option. These two conditions are stated as

$$\frac{1/2}{1-\delta} < V_O < \frac{p_n}{1-\delta}. \tag{1}$$

Before proceeding to the players' strategies, I want to discuss some modeling choices. I assume that the recommendation itself is payoff relevant, i.e. E receives his bonus whenever he recommends the bonus option and C receives his payoff whenever the recommendation fits his need. Put differently, there is no real decision by C whether or not to follow the recommendation. This is not unreasonable because C has uniform beliefs and therefore he cannot draw any inference about the likelihood that the recommendation fits his needs from the recommendation itself. Given that C continued in the previous period, and thereby asked for more advice, following this advice seems to be logical. It is assumed that at the end of a period both C and E observe whether the given recommendation fitted the consumer's needs. In the examples mentioned earlier,

[3]The finiteness of P simplifies the exposition but does not affect the results. As p_i cannot increase above 1, learning has to eventually flatten out in the sense that precision has to converge to some upper bound as i gets large. Finiteness of P relieves us from the notational burdensome task of taking limits in certain proofs and allows to use backwards induction right away.

this last assumption is reasonable: A sales person observes whether the consumer tries to return the product, the civil servant observes whether his draft is pushed forward and the doctor will find out whether the patient recovers.[4] In the following, I will use the word *hit* (*miss*) to refer to the event that the recommendation fitted (did not fit) the consumer's needs.

This is repeated each period as long as C does not end the relationship by choosing his outside option. Note that the model assumes independence at several points. First, the bonus option is independent from the consumer's needs. This is one of the main differences to the cheap talk literature and appears naturally in the exmaples of the introduction. Second, there is a certain independence across time in the sense that the consumer's needs and the bonus options are drawn independently each period. One way to interpret this is that the requests of the consumer are unrelated, e.g. searching for an Italian restaurant in one period and for news in another period in the search engine example or suffering from different diseases in the patient doctor example. In the financial advice example, the market environment and the set of available products may change from period to period.

As argued before, E gets to know the consumer better and therefore the precision of E's signal should increase over time. Depending on the application, the precision might increase either after each interaction or after each hit or not at all. It seems realistic that a fitting recommendation tells more about a consumer's preferences than a non-fitting one. The assumption made here is that the precision is increasing in the number of past hits and this increase is deterministic and commonly known by C and E. That is, no learning happens after misses. The special case of no learning is analyzed as a starting point later on.

It is worthwhile to note that no meaningful advice would be possible if the game was not infinitely repeated. Consider the static case. E has no incentives to recommend anything but his bonus option. C will therefore not get information about which option is more likely to fit his needs. A similar situation emerges in a finitely repeated game. The static analysis applies to the last period. Given that no meaningful communication takes place in the last period, E should end the game after the second but last period (regardless of history). Anticipating this, E will optimally recommend his bonus option in the second but last period no matter what his signal is. Iterating this reasoning the game unravels and no meaningful advice in any period is possible. In the infinitely repeated game the situation changes because future bonusses can potentially motivate E to give truthful advice even if his bonus option is option 2. As there is no last period,

[4] In the search engine example, the search engine observes whether the link was clicked and – in the case of Google – to the extent that the target website uses GoogleAnalytics, csi.gstatic, GoogleAdSense or a GooglePlus button, Google will also receive information on the consecutive behavior of the user on the target website.

there is no period where these dynamic incentives break down.

What are the *strategies* of the players in this game? I assume that the players condition their decision only on observed, payoff relevant information. That is, C's choice depends only on the sequence of hits and misses in previous periods.[5] E has to decide in each period which option to recommend. His decision depends on his posterior belief, his bonus action and the history of hits and misses. In principle, his decision could also depend on the history of bonus options but this possibility is neglected because his current and future payoffs do not depend on this information (neither directly nor indirectly as C's strategy cannot condition on this information which C did not observe).

In the following, I employ two often used notions of equilibrium and compare their outcomes. Both put further restrictions on strategies. First, *Markov equilibrium* where strategies condition only on actions and information of the present period and a payoff relevant state variable. The state variable is the current precision p_k. Consequently, E's strategy is a function $P \times \{1, 2\} \to [0, 1]$ which assigns to every $p_k \in P$ and the identity of the bonus option a probability of recommending option 1. C's strategy is a function $P \times \{hit, miss\} \to [0, 1]$ which assigns to every $p_k \in P$ and the event that this periods recommendation fitted C's needs a probability of continuing.

The second notion of equilibrium is *grim trigger*. C continues as long as the recommendation are hits. He ends the game when a recommendation turns out to be a miss. E plays a best response to this strategy. (Of course, it remains to show that C's grim trigger strategy is a best response to E's best response but this turns out to be straightforward unless V_O is too high.)

3. Analysis

3.1. Markov equilibrium

Note first that "babbling Markov equilibrium" always exists. In this equilibrium E will always recommend his bonus option and C will always stop the game. Clearly, these are mutually best responses given assumption (1). The interesting question is therefore not whether a Markov equilibrium exists but whether a Markov equilibrium with some information transmission exists. Before answering this questio in general, it is useful to analyze the case without learning where the precision of E's signal stays constant. If C does not stop the game beforehand, this situation occurs after n hits when E's signal has precision p_n.

[5]In principle, C observes the specific recommendations but since the specific labels are not observed by him, he is unable to condition his strategy on these labels.

3.1.1. Model without learning

Without learning the state never changes and therefore a "Markov strategy" will only condition on this periods information/actions. That is, a strategy for E consists of two probabilities giving the probability of recommending option 1 in case of (i) it is the bonus option and (ii) it is not. Similarly, a strategy of C consists of two probabilities of continuing: one in case of a hit and one in case of a miss.

The probability of continuing is in equilibrium (weakly) higher in case of a hit than in case of a miss. Otherwise, E would have incentives to give worst possible advice, i.e. always recommend option 2 when it is the bonus option (and possibly even when it is not) which automatically implies, by (1), that C is better off ending the game.

As the probability of continuing is higher in case of a hit than in case of a miss, it is optimal for E to recommend option 1 if option 1 is the bonus option. In this case the incentives of C and E are aligned. E's strategy can therefore be reduced to one probability α with which he recommends option 1 in case option 2 is the bonus option.

While other equilibria can exist, I will concentrate on the case where C continues with probability 1 in case of a hit. Note that this incentivizes E most to be truthful. The restriction is not problematic: It is not hard to show that whenever a non-babbling Markov equilibrium exists, there exists a Markov equilibrium in which C continues with probability 1 in case of a hit. Furthermore, this is the equilibrium Pareto dominates all other Markov equilibria. Given this restriction, C's strategy is simply a probability β of continuing in case the recommendation is a miss.

Denote E's equilibrium value, i.e. his discounted expected payoff stream at the start of a period (even before knowing the identity of the bonus option), by Π. If option 2 is the bonus option, E prefers recommending option 1 if

$$\begin{aligned} p\delta\Pi + (1-p)\beta\delta\Pi &\geq 1 + p\beta\delta\Pi + (1-p)\delta\Pi \\ \Leftrightarrow \beta &\leq \frac{(2p-1)\delta\Pi - 1}{(2p-1)\delta\Pi}. \end{aligned} \tag{2}$$

Denote C's equilibrium value by V and note that C is willing to continue only if $V \geq V_O$. As this is independent of whether the current periods recommendation was a hit or a miss and given that C continues for sure after a hit, C either has to continue with probability 1 after a miss as well, $\beta = 1$, or C has to be indifferent, $V = V_O$. The former cannot happen in equilibrium: (2) cannot hold for $\beta = 1$ and E would therefore always recommedn his bonus option. By (1), C would then, however, strictly prefer not to continue. Therefore, $V = V_O$ in equilibrium and consequently α has to be such that[6]

$$\begin{aligned} V_O &= \frac{1}{2}p + \frac{1}{2}\left(\alpha p + (1-\alpha)(1-p)\right) + \delta V_O \\ \Leftrightarrow \alpha &= \frac{2(1-\delta)V_O - 1}{2p-1}. \end{aligned} \tag{3}$$

By (1), $\alpha \in (0,1)$. Hence, E uses a mixed strategy in an information equilibrium and E is only willing to mix if (2) holds with equality. Given these equilibrium strategies one can determine the values and obtain conditions for the existence of a non-babbling Markov equilibrium.

Proposition 1. *An non-babbling Markov equilibrium exists in the model without learning if and only if*

$$\frac{1-\delta}{\delta} \leq \frac{4p-3}{2}. \tag{4}$$

In such an equilibrium $V = V_O$ and $\Pi > 0$ and in the Pareto optimal Markov equilibrium α is given by (3) and $\beta = 1 - 1/[(2p-1)\delta\Pi]$.

Proof of proposition 1: With $\beta = 1 - 1/[(2p-1)\delta\Pi]$, it is straightforward to determine Π:[7]

$$\begin{aligned} \Pi &= \frac{1}{2} + \left(p + \beta(1-p)\right)\delta\Pi = \frac{1}{2} + \left(p + (1-p) - \frac{1-p}{(2p-1)\delta\Pi}\right)\delta\Pi \\ \Leftrightarrow \Pi &= \frac{4p-3}{(1-\delta)(4p-2)}. \end{aligned}$$

Plugging this back into (2) (with equality) yields

$$\beta^* = 1 - \frac{2(1-\delta)}{\delta(4p-3)}.$$

A non-babbling Markov equilibrium exists if $\beta^* \in [0,1]$ which is the case if

$$\frac{1-\delta}{\delta} \leq \frac{4p-3}{2}.$$

□

3.1.2. Model with learning

Again it is straightforward to see that E will always recommedn option 1 if option 1 is the bonus option. Furthermore, I will concentrate on on-babbling Markov equilibria in

[6]As C is indifferent, we can determine his value $V = V_O$ by writing down the expected payoff stream if he continued for sure this period.

[7]As E is indifferent in case option 2 is his bonus option, his value is as if he recommended always option 1.

which C continues for sure in case of a hit. Strategies are therefore given by α^k and β^k for $k \in \{1, 2, \dots n\}$. The players' values, i.e. their expected discounted payoff streams at the start of a period with precision p^k, will be denoted by Π^k and V^k. It is clear from the previous subsection that such an equilibrium can only exist if (4) holds (for $p = p_n$). While this condition is necessary, it is not sufficient for the existence of a non-babbling Markov equilibrium and will therefore be generalized below.

The first step is to show that in no period E will recommend option 1 regardless of the identity of the bonus option while C continues regardless of whether the recommendation is a hit or a miss. While this property is not surprising it is also not totally straightforward: After all recommending option 1 gives E a higher chance to move to the next highest precision and in principle it would be possible that this motivates him to be truthful (if Π^{k+1} is sufficiently larger than Π^k).

Lemma 1. *In Markov equilibrium $\alpha^k = \beta^k = 1$ cannot hold for any k because E's best response to $\beta^k = 1$ is $\alpha^k = 0$.*

Proof of lemma 1: Suppose $\beta^k = 1$ and consider the case when option 2 is E's bonus option. E's value as a function of α is then

$$\begin{aligned}
\Pi^k &= \frac{1}{2}\left(p^k\delta\Pi^{k+1} + (1-p^k)\delta\Pi^k\right)(1+\alpha) + \frac{1}{2} + \frac{1}{2}(1-\alpha)\left(p^k\delta\Pi^k + (1-p^k)\delta\Pi^{k+1} + 1\right) \\
&= \frac{1}{2}(2-\alpha) + \frac{1}{2}\delta\Pi^{k+1}(2p^k\alpha + 1 - \alpha) + \frac{1}{2}\delta\Pi^k(1+\alpha-2p^k\alpha) \\
\Leftrightarrow \Pi^k &= \frac{2-\alpha}{2-\delta-\delta\alpha+2p^k\delta\alpha} + \frac{\delta-\delta\alpha+2p^k\delta\alpha}{2-\delta-\delta\alpha+2p^k\delta\alpha}\Pi^{k+1}.
\end{aligned}$$

This implies

$$\begin{aligned}
\Pi^k_{\alpha=0} &= \frac{2}{2-\delta} + \frac{\delta}{2-\delta}\Pi^{k+1} \\
\Pi^k_{\alpha=1} &= \frac{1}{2(1-\delta+p^k\delta)} + \frac{2p^k\delta}{2(1-\delta+p^k\delta)}\Pi^{k+1}
\end{aligned}$$

where $\Pi^k_{\alpha=1}$ is E's equilibrium value in the supposed equilibrium (where E uses the strategy $\alpha^k = 1$) and $\Pi^k_{\alpha=0}$ is a deviation value that E would obtain if he deviated from the supposed equilibrium strategy by choosing $\alpha^k = 0$ (without changing his strategy for $k' \neq k$). For $\alpha^k = 1$ to be optimal $\Pi^k_{\alpha=1} \geq \Pi^k_{\alpha=0}$ has to hold. However, it is now shown that $\Pi^k_{\alpha=0} > \Pi^k_\alpha$ for any $\alpha > 0$. This inequality can be written as

$$\begin{aligned}
\frac{2}{2-\delta} + \frac{\delta}{2-\delta}\Pi^{k+1} &> \frac{2-\alpha}{2-\delta-\delta\alpha+2p^k\delta\alpha} + \frac{\delta-\delta\alpha+2p^k\delta\alpha}{2-\delta-\delta\alpha+2p^k\delta\alpha}\Pi^{k+1} \\
\Leftrightarrow 4-2\delta-2\delta\alpha+4p^k\alpha\delta+(2-\delta-\delta\alpha+2p^k\alpha\delta)\delta\Pi^{k+1} & \\
&> 4+\alpha\delta-2\delta-2\alpha+(2+\alpha\delta-\delta-2\alpha+4p^k\alpha-2p^k\alpha\alpha)\delta\Pi^{k+1} \\
\Leftrightarrow -\alpha\delta+2\alpha+4\alpha p^k\delta &> (1-\delta)(4p^k\alpha-2\alpha)\delta\Pi^{k+1}.
\end{aligned}$$

The latter inequality is true for all $\alpha > 0$ because Π^{k+1} is bounded from above by $1/(1-\delta)$ (which would be E's discounted payoff stream if he always recommended his bonus option and C always continued) and the previous inequality holds with $1/(1-\delta)$ in place of Π^{k+1}:

$$\begin{aligned}
-\alpha\delta+2\alpha+4\alpha p^k\delta &> (1-\delta)(4p^k\alpha-2\alpha)\frac{\delta}{1-\delta} \\
\Leftrightarrow \alpha(2+\delta) &> 0.
\end{aligned}$$

This shows that $\alpha^k = 0$ is the only best response to $\beta^k = 1$ andtherefore $\Pi^k_{\alpha=1} < \Pi^k_{\alpha=0}$. Consequently, $\beta^k = \alpha^k = 1$ cannot be an equilibrium. □

Lemma 2. *In every Markov equilibrium $V^k = V_O$ for all $k \in \{1, 2, \ldots, n\}$.*

Proof of lemma 2: Proposition 1 implies $V^n = V_O$. Suppose $V^k > V_O$ for some k and let k' be the highest such k. Then $\alpha^{k'}$ must be sufficiently high yield a higher payoff than $(1-\delta)V_O$ to C in every period with precision $p^{k'}$. Now consider C's decision problem after a miss in a period with precision $p^{k'}$. As $V^{k'} > V_O$ by the definition of k', C strictly prefers to continue. Hence, $\beta^{k'} = 1$. But E's best response to $\beta^{k'} = 1$ is $\alpha^{k'} = 0$, see the proof of lemma 1. But given that $V^k = V_O$ for all $k > k'$ by the definition of k' and given that $\alpha^{k'} = 0$ clearly $V^{k'} < V_O$ contradicting the definition of k'. □

Lemma 2 implies E's strategy in Markov equilibrium. If the game reaches precision p^k with positive probability in a Markov equilibrium, then E has to mix such that C is indifferent between continuing and stopping. That is,

$$\begin{aligned}
V_O &= \frac{1}{2}p^k + \frac{1}{2}\left(\alpha^k p^k + (1-\alpha^k)(1-p^k)\right) + \delta V_O \\
\Leftrightarrow \alpha^k &= \frac{2(1-\delta)V_O - 1}{2p^k-1}. \qquad (5)
\end{aligned}$$

Note that α^k as given by (5) is in $(0,1)$ by assumption (1). Consequently, E must be indifferent between recommending either option if the bonus option is option 2. This indifference condition determines β^k:

$$\begin{aligned} 1 + p^k\beta^k\delta\Pi^k + (1-p^k)\delta\Pi^{k+1} &= 0 + p^k\delta\Pi^{k+1} + (1-p^k)\beta^k\delta\Pi^k \\ \Leftrightarrow \beta^k &= \frac{(2p^k-1)\delta\Pi^{k+1}-1}{(2p^k-1)\delta\Pi^k}. \end{aligned} \tag{6}$$

Note that Π^n is given by the stationary equilibrium value derived in the proof of proposition 1. Starting from there Π^{n-1} and β^{n-1} can be obtained and through backwards induction all other β^k and Π^k can be obtained as well. A non-babbling Markov equilibrium will exist if all such obtained β^k are in $[0,1]$. The following proposition gives a necessary and sufficient condition for exactly this.

Proposition 2. *A non-babbling Markov equilibrium exists if and only if*

$$\frac{\delta^{n-1}}{1-\delta}\frac{4p^n-3}{4p^n-2} + \sum_{k=0}^{n-2}\delta^k\frac{4p^k-3}{4p^k-2} \geq \frac{1}{\delta(2p^0-1)}. \tag{7}$$

In this Markov equilibrium, $V^k = V_O$ *and*

$$\Pi^k = \frac{\delta^{n-k}}{1-\delta}\frac{4p^n-3}{4p^n-2} + \sum_{j=0}^{n-k-1}\delta^j\frac{4p^j-3}{4p^j-2} \tag{8}$$

for $k \in \{1,2,\ldots,n\}$ *and* α^k *and* β^k *are given by (5) and (6) respectively.*

Proof of proposition 2: As E is mixing in a non-babbling Markov equilibrium when the bonus option is option 2, his value will equal the value he would get if he always recommended option 1 (keeping C's strategy fix):

$$\Pi^k = \frac{1}{2} + p^k\delta\Pi^{k+1} + (1-p^k)\beta^k\delta\Pi^k.$$

Plugging (6) in for β^k yields

$$\begin{aligned} \Pi^k &= \frac{1}{2} + p^k\delta\Pi^{k+1} + (1-p^k)\delta\Pi^{k+1} - \frac{1-p^k}{2p^k-1} \\ \Leftrightarrow \Pi^k &= \delta\Pi^{k+1} + \frac{4p^k-3}{4p^k-2}. \end{aligned}$$

Recall from proposition 1 that $\Pi^n = (4p^n-3)/[(4p^n-2)(1-\delta)]$. Using this as a starting point for backwards induction in the previous equation yields (8).

Next I will show that Π^k is strictly increasing in k. Let $h(p^k) = (4p^k-3)/(4p^k-2)$

and note that $h' > 0$ for $p^k \in [0,1]$. This allows to write

$$\begin{aligned}
\Pi^{k+1} - \Pi^k &= (1-\delta)\Pi^{k+1} - h(p^k) \\
&= \frac{\delta^{n-k-1}}{1-\delta} h(p^n) - \frac{(1-\delta)h(p^k)}{1-\delta} + (1-\delta)\sum_{j=0}^{n-k-2} \delta^j h(p^{j+k+1}) \\
&= \frac{\delta^{n-k-1}}{1-\delta}\big(h(p^n) - h(p^k)(1-\delta)\big) + (1-\delta)\sum_{j=0}^{n-k-2} \delta^j \big(h(p^{j+k+1}) - h(p^k)\big) > 0
\end{aligned}$$

where the inequality holds because p^k is strictly increasing in k and h is strictly increasing in p^k. Therefore, $\Pi^{k+1} - \Pi^k > 0$, i.e. Π^k is strictly increasing in k.

For existence of a non-babbling Markov equilibrium, a $\beta^k \in [0,1]$ has to exist to make E indifferent between the two recommendations in case option 2 is the bonus option. For $\beta^k = 1$, E strictly prefers to recommend option 1. As the incentives to recommend option 1 are strictly decreasing in β^k, a $\beta^k \in [0,1]$ will exist if and only if E prefers recommending option 2 (in case it is the bonus option) for $\beta^k = 0$. That is, if

$$\begin{aligned}
1 + (1-p^k)\delta\Pi^{k+1} &\leq p^k\delta\Pi^{k+1} \\
\Pi^{k+1} &\geq \frac{1}{\delta(2p^k - 1)}.
\end{aligned}$$

This condition is most demanding for $k = 0$ because p^k and Π^k are both increasing in k. Hence, a non-babbling Markov equilibrium exists if and only if $\Pi^1 \geq 1/(\delta(2p^0 - 1))$. Plugging in the above derived expression for Π^1, this is condition (7). □

3.2. Simple grim trigger strategies and m-equilibrium

Like most repeated games, the game described here has multiple perfect Bayesian Nash equilibria. I will now concentrate on a class of equilibria in which C uses the following particularly simple strategy: C continues the relationship unless the past $m \geq 1$ recommendations were misses. After m consecutive misses, C stops the game and consumes his outside option. As the strategy resembles somewhat the grim trigger strategies taught in introductory game theory, I will call this strategy a simple grim trigger strategy of length m or *m-strategy* for short. A perfect Bayesian Nash equilibrium in which C uses an m-strategy is called *m-equilibrium*.

When can a m-strategy be optimal for C? First, C must have a continuation value of at least V_O after every history that includes less than m consecutive misses. Second, continuing after m misses must lead to a continuation value of at most V_O. The latter can be achieved easily: It is, by (1), optimal to end the game if E recommends his bonus option in all following periods. In an m-equilibrium, continuing after m or more misses is clearly off the equilibrium path. Hence, the following off path beliefs of E will make this response optimal: If C has continued after m misses before, then E believes that

C will end the game in the next period no matter whether there is a miss or hit in the current period. Given this belief, it is clearly optimal to recommend the bonus option now. This implies that it is indeed optimal for C to end the game after m (or more) misses. These off path beliefs are not ruled out by perfect Bayesian Nash equilibrium or normal refinements.

Given this off path construction, the following steps are sufficient to construct an m-equilibrium. First, derive E's best response to C's m-strategy. Second, verify that C's continuation value on the equilibrium path is at least V_O.

4. Results

What is E's best response to an m-strategy? In a given period, E is always tempted to recommend the bonus option to ensure himself a payoff of 1. The downside of this choice is that a miss is quite likely if the posterior belief that the bonus action fits C's needs is low. An additional miss brings E closer to the end of the relationship which stops the bonus stream forever and therefore leads to a payoff of zero for E. It is immediate that E will always recommend option 1 if option 1 is the bonus option.

I denote the value of the expected discounted bonus stream after t consecutive misses by $\Pi^t(q)$. After $t-1$ consecutive misses, it is optimal for E to recommend the bonus option r_k instead of r_1 if

$$p_k(q_l)\delta\Pi^0(q_{l+1}) + (1-p_k(q_l))\delta\Pi^t(q_l) + 1 \geq p_1(q_l)\delta\Pi^0(q_{l+1}) + (1-p_1(q_l))\delta\Pi^t(q_l) \quad (9)$$

where q_l is the current quality. If this inequality holds for r_k being the bonus option, then it will also hold if r_j with $j < k$ is the bonus option (recall that we ordered option according to E's posterior). Hence, E's best response takes the form of a cutoff $k^t(q_l)$: If the bonus option is r_j with $j \leq k^t(q_l)$, E recommends the bonus option and E recommends r_1 otherwise.

Lemma 3. *The cutoff $k^t(q_l)$ after $t-1$ consecutive misses is*

- *1 iff $p_2(q_l) < p_1(q_l) + A^t(q_l)$*
- *n iff $p_n(q_l) \geq p_1(q_l) + A^t(q_l)$*
- *k where $p_{k+1}(q_l) \leq p_1(q_l) + A^t(q_l) < p_k(q_l)$ else*

where $A^t(q_l) = -1/(\delta\Pi^0(q_{l+1}) - \delta\Pi^t(q_l))$ and q_l is the current quality level.

Clearly, the previous lemma is only of limited use as long as E's value function Π is not known. However, the current formulation is already enough to derive some intuitive intermediate results. The first one states that E's value is decreasing in the number of misses. Intuitively, more misses mean that the risk of getting to m consecutive misses – and therefore the end of the relationship – is higher.

Lemma 4. *Holding q_l fixed, $\Pi^t(q_l)$ is decreasing in t.*

The second result states that the cutoff $k^t(q_l)$ is decreasing in the number of misses. That is, C gets better information the higher the number of consecutive misses. The intuition is that E is getting more and more desperate for a hit as an additional miss brings him dangerously close to m consecutive misses which lead to the end of the relationship.

Lemma 5. *The cutoff $k^t(q_l)$ is decreasing in t for a fixed q_l.*

4.1. Equilibrium without learning

It proves useful to first consider the case where there is no learning. This means that q is the same in all periods and the dependence of p_k, k^t and Π^t on q can simply be dropped here. To derive E's best response, it is necessary to determine E's value function Π^t for $t = 0, \ldots, m-1$. As soon as these values are derived, lemma 3 gives E's strategy.

Note that Π^{t-1} can – given the optimal cutoffs from lemma 3 – be written as

$$\begin{aligned}
\Pi^{t-1} &= \sum_{j=1}^{k^t} \frac{1}{n}\left[1 + p_j\delta\Pi^0 + (1-p_j)\delta\Pi^t\right] + \sum_{j=k^t+1}^{n} \frac{1}{n}\left[p_1\delta\Pi^0 + (1-p_1)\delta\Pi^t\right] \qquad (10) \\
&= \frac{k^t}{n} + \frac{\delta\Pi^0}{n}\left(p_1\left(n-k^t\right) + \sum_{j=1}^{k^t} p_j\right) + \frac{\delta\Pi^t}{n}\left((1-p_1)(n-k^t) + \sum_{j=1}^{k^t}(1-p_j)\right).
\end{aligned}$$

As $\Pi^m = 0$, (10) allows to write Π^{m-1} as a function of Π^0. That is, one can write $\Pi^{m-1} = \tilde{\Pi}^{m-1}(\Pi^0)$ where $\tilde{\Pi}$ simply uses (10). Given $\tilde{\Pi}^{m-1}(\Pi^0)$, (10) can be used again to derive $\Pi^{m-2} = \tilde{\Pi}^{m-2}(\Pi^0)$. Iterating this procedure, one eventually arrives at $\Pi^0 = \tilde{\Pi}^0(\Pi^0)$. Any fixed point of $\tilde{\Pi}^0$ gives then E's value function in a best response to C's m-strategy. As shown in the proof of the following proposition, $\tilde{\Pi}^0$ has a unique fixed point. Hence, there is a unique, well defined best response of E to C's m-strategy. Given this best response, an m-equilibrium exists if C's continuation value is at least V^0 after any history.

Proposition 3. *Take $m \in \mathbb{N}$. An m-equilibrium exists if the outside option is sufficiently small (but above $1/(n - n\delta)$).*

4.2. Equilibrium with learning

In the previous subsection, E's value function was derived using backwards induction starting from $\Pi^m = 0$ and a fixed point. While the backwards induction approach is still valid with learning, the fixed point argument is not. The reason is that Π^0 will now depend on q. That is, Π^0 after l hits is not the same as Π^0 after $l+1$ hits because quality increases in the number of hits.

However, there is a way to circumvent this problem. With learning, E's posterior $(p_1(q_l), \dots, p_n(q_l))$ depends on the number of hits. As the number of hits grows to infinity, E's posterior converges to a limit distribution $(p_1^*, \dots, p_n^*)$. For this limit distribution, an equilibrium without learning can be constructed as in the previous subsection. In particular, a value Π^{0*} can be obtained. As $\Pi^0(q_l)$ will converge to Π^{0*} as the number of hits grows large, the error made by using Π^{0*} instead of $\Pi^0(q_l)$ after L hits is arbitrarily small if L is chosen large enough. By using $\Pi^0(q_l) \approx \Pi^{0*}$ for $l \geq L$, it is possible to derive $\Pi^t(q_l)$ for $l < L$ using backwards induction. The following proposition states that one can get arbitrarily close to an m-equilibrium using the method described here.

Proposition 4. *For any $\varepsilon > 0$, there is a $L \in \mathbb{N}$ such that the error for each $\Pi^t(q_l)$ is less than ε.*

It is even possible to construct the exact equilibrium in generic cases. Assume that in the limit m-equilibrium, i.e. the equilibrium without learning where E has the limit posterior $(p_1^*, \dots, p_n^*)$, A^t is such that E is not indifferent between two cutoffs (for all $t \in \{1, \dots, m-1\}$). This will be the case for generic parameter values and in this case the exact equilibrium can be computed: In this case, E will find the same cutoffs as in the limit optimal if the number of hits is greater than some L. This allows to write down the expected profits of E (for more than L hits) as an infinite series as his optimal strategy is clear for all consecutive periods. Given $\Pi^0(q_L)$, $\Pi^t(q_l)$ for $l \leq L$ can then be derived exactly by backwards induction.

4.3. Welfare dynamics

The most interesting question in this model is who benefits from the learning and why. To state the results, an additional condition is needed. I will say that a signal technology satisfies *supermodularity* if the posterior has increasing differences, that is if $p_k(q) - p_j(q)$ is increasing in q if and only if $k \leq j$. A weaker condition is to require increasing differences only with respect to the most likely option:

Condition 1 (SM1)**.** *A signal technology satisfies SM1 if $p_1(q) - p_j(q)$ is increasing in q for every $j \in \{2, \dots, n\}$.*

The following proposition states that both – C and E – will benefit from increasing quality if the signal technology satisfies SM1. That is, E cannot capture all rents associated with his learning in any m-equilibrium. To state the result, denote the consumer's expected discounted benefit after t consecutive misses at quality q_l by $V^t(q_l)$ (in a given m-equilibrium).

Proposition 5. *In an m-equilibrium, $\Pi^t(q_l)$ is strictly increasing in q_l. If SM1 holds, then $V^t(q_l)$ is also strictly increasing in q_l and the optimal cutoff levels $k^t(q_l)$ are weakly decreasing in q_l.*

It is relatively intuitive that E's surplus is increasing in quality. Even if E did not adjust the optimal cutoffs k^t, he would still benefit from having a more accurate signal which implies that the option fitting C's needs is more likely to be among the first k^t options.

Why is consumer surplus increasing in quality? Since E's surplus is increasing in quality, E is relatively more interested in interacting with C in the future if quality is higher. E is therefore willing to give better information to C in order to ensure that C does not end the relationship. In short, C is a more valuable customer for E and therefore he will try to make him happy in order to keep him.

A direct implication of this is that C prefers a setting with personalization to a setting without personalization. If C interacted each period with a different expert, or if E was not able/allowed to keep any records of prior interactions with C, then no learning would occur. As consumer surplus is strictly increasing in q_l, this would, however, harm C.

What is the role of SM1 and why is it necessary for the results on consumer surplus? In a given period, say after t consecutive misses, expected consumer surplus in this period can be written as

$$\frac{n - k^t(q)}{n} p_1(q) + \frac{1}{n} \sum_{j=1}^{k^t(q)} p_j(q).$$

This expected consumer surplus increases in q, if $k^t(q)$ is decreasing in q: Both $p_1(q)$ and $\sum_{j=1}^{k^t(q)} p_j(q)$ are increasing in q because the posterior corresponding to higher q is first order stochastically dominated. If $k_t(q)$ strictly decreases, then a $p_j(q)$ is substituted by $p_1(q) > p_j(q)$.

Condition SM1 ensures that $k^t(q)$ is decreasing in q. To see this, one has to return to lemma 3. It is easy to show that $A^t(q)$ is decreasing in q. Condition SM1 ensures that $p_1(q) - p_k(q)$ is increasing. Taking these two observations together, lemma 3 implies that the optimal cutoff is decreasing in q.

This leads to the interesting questions under which circumstances C can lose from E's learning. Clearly, C can only lose if $k^t(q)$ increases in q. From lemma 3, it follows that this could be the case if $p_1(q) - p_k(q)$ was decreasing in q. This is the case if the learning takes the form of being able to exclude unlikely options with higher certainty. Put differently, C can suffer from E's learning if higher q shifts probability mass in the posterior from the unlikely options n, $n-1, \ldots$ to options around $k^t(q)$. Learning will benefit C if higher quality means that probability mass in the posterior is shifted mainly towards the most likely option (this is SM1).

Note that $k^t(q)$ cannot increase all the time because there are only n options. Furthermore, $p_1(q) - p_n(q)$ is increasing in q as higher quality leads to a first order stochastically dominated posterior. Therefore, $k^t(q) < n$ implies $k^t(q') < n$ for all $q' > q$. In a nutshell, even if SM1 is not satisfied, C will benefit from E's learning eventually in the sense that $V^t(q)$ attains its global minimum at a certain quality and achieves strictly higher values for all higher quality levels. Nevertheless, it is possible that expected consumer surplus is higher initially than in the limit, i.e. $V^t(q_0) > V^t(q^*)$ cannot be ruled out if SM1 is violated.

(CS maximizing m-equilibrium, comparative static in bonus)

5. Discussion

The results in the previous section have implications for anonymization. Activists and experts alike recommend measures to preserve anonymity online. Although many of these recommendations are easy to follow, e.g. using an anonymized version of Google instead of Google itself, hardly any internet user follows them. The analysis above indicates that consumers might be right when not anonymizing: Personalized recommendations are more valuable not only from a total surplus perspective but also from a consumer perspective (at least if SM1 is satisfied). The reason is simple. The more past usage data is available, the more valuable the customer is. The expert, e.g. Google, does not want to risk loosing valuable customers. Hence, a customer enjoys better service the more past usage data the expert has.

The same principle applies in other applications and explains why long term advisers are more valuable than short term advisers. The model gives an interesting prediction for the hazard rate, i.e. the probability that a consumer ends the relationship after a certain number of hits given that he has not ended it yet. Given SM1, the consumer gets better advice over time in an m-equilibrium; see proposition 5. This means that the probability of m consecutive misses after a hit is smaller if q_l is higher. Put differently, the hazard rate is – in this sense – decreasing over time.

Of course, there are some caveats to these results. The first is that the outside option of the consumer was kept constant. If the outside option is an alternative expert, this could change. To give an example, say there are two experts and everyone agrees that expert 1 is slightly more knowledgeable than expert 2. The outside option corresponds then to getting advice from expert 2. If everyone uses expert 1, however, expert 2 might be out of business and take up a different job. In the long term, the outside option could therefore decline and might eventually drop below $1/(n - n\delta)$. In this case, the unique equilibrium is that the expert recommends his bonus action every period and consumers would suffer. However, an m-equilibrium is not sensitive to lower outside options as long as the outside option remains above $1/(n - n\delta)$.

Another caveat, in particular in connection with anonymization of online activities, is that the model did not deal with possible extortion resulting from abuse of data outside the advice relationship. According to the model, a customer benefits from personalized advice and a prerequisite for such personalized advice is that data on past interactions is kept. If this data finds its way in the hands of a third party, it could in some cases be used against the consumer by this third party; think of health or financial records. Such extortion from third parties is beyond the scope of this paper.

A more interesting question is whether there are equilibria in the model where the implications do not hold, i.e. is there a Bayesian Nash equilibrium (which necessarily cannot be an m-equilibrium) such that the equilibrium consumer surplus equals his outside option. Indeed there are equilibria that give the consumer a rather small consumer surplus: Let $\bar{m}$ be the highest m such that the consumer has a consumer surplus above his outside option in this m-equilibrium. Now consider the following consumer strategy: C is willing to tolerate $\bar{m}+1$ consecutive misses before ending the relationship initially. After the first hit, C returns to his $\bar{m}$-strategy. Let E play the best response to this consumer strategy.[8] Clearly, consumer surplus will be lower than in the $\bar{m}$ equilibrium. If consumer surplus is still above the outside option, one can try to construct an equilibrium where C plays the $\bar{m}+1$ strategy until two hits occur etc. If consumer surplus is already below the outside option, one can try an alternative strategy where C allows for $\bar{m}+1$ consecutive misses not initially but after l hits (and E plays a best response to this strategy). As the consumer surplus increases in q_l, increasing l implies that $V^0(q_l)$ is higher and therefore consumer surplus might not drop below the outside option when changing C's strategy as described. If there is no such l, then consumer surplus is already quite close to the outside option and does not increase substantially in the number of hits.

Using similar changes in C's strategy repeatedly can lead to equilibria where C does not benefit from E's learning (in the following let E always play his respective best response): Start with the $\bar{m}$-strategy. Then adapt C's strategy by changing to the $\bar{m}+1$-strategy as soon as q_l is sufficiently high such that $V^0(q_l)$ in the $\bar{m}+1$ equilibrium is above C's outside option. Let the consumer switch to the $\bar{m}+2$ strategy as soon as q_l is sufficiently high such that $V^0(q_l)$ is above C's outside option in the $\bar{m}+2$ equilibrium etc. This equilibrium does not support the claim of the paper that C benefits from E's learning. However, C plays quite a complicated strategy here (and this has to be common knowledge among C and E in equilibrium) while C's strategy in an m-equilibrium is relatively simple and seems in line with the phrase "everyone

[8]This best response can be derived via backwards induction as before.

deserves a second chance" (which is often understood to imply "…but not a third"). Also having a consumer playing a complicated strategy though he would fare better in a simple equilibrium appears somewhat counterintuitive. In total, it seems unlikely that the complicated adaptive equilibrium has practical relevance for consumer markets.

This discussion illustrates that the techniques used to derive m equilibria can also be used to derive equilibria in which C does not use an m-strategy, i.e. strategies where the number of consecutive misses before C ends the relationship depends on the number of prior hits. As long as the number of different m_ls used by C is finite, E's best response can be derived via backwards induction from the limit equilibrium as before.

6. Conclusion

Appendix

Proofs

Proof of lemma 3: The optimal cutoff is determined by the two inequalities

$$p_k(q_l)\delta\Pi^0(q_l) + (1 - p_k(q_l))\delta\Pi^t(q_{l+1}) + 1 \;\geq\; p_1 q_l \delta\Pi^0 q_{l+1} + (1 - p_1 q_l)\delta\Pi^t q_l \tag{11}$$

$$p_{k+1}(q_l)\delta\Pi^0(q_{l+1}) + (1 - p_{k+1}(q_l))\delta\Pi^t(q_l) + 1 \;\leq\; p_1(q_l)\delta\Pi^0(q_{l+1}) + (1 - p_1(q_l))\delta\Pi^t(q_l) \tag{12}$$

which state that recommending r_k is optimal if r_k is the bonus alternative and recommending r_1 is optimal if r_{k+1} is the bonus alternative. Rearranging gives the expression in the lemma. □

Proof of lemma 4: The argument is a simple strategy stealing argument. The best response cutoffs are $k^t(q_l)$. I will compare $\Pi^t(q_l)$ and $\Pi^{t+1}(q_l)$ for some arbitrary $t \in \{0, \dots, m-1\}$. Suppose E used after t misses the cutoff $g^t = k^{t+1}(q_l)$ and suppose further that E used the cutoffs $g^s = k^{s+1}(q_l)$ after $s \in \{t+1, m-2\}$ misses and $g^{m-1} = n$. In this case, his expected payoff would be $\hat{\Pi}^t(q_l) = \Pi^{t+1}(q_l) + 1 * \alpha > \Pi^{t+1}(q_l)$ where $\alpha > 0$ is the probability of misses in all periods from t up to m. Using the optimal cutoffs $k^t(q_L)$ instead of g^t, can only increase $\Pi^t(q_l)$ further above $\hat{\Pi}^t(q_l)$. Hence, $\Pi^t(q_l) \geq \hat{\Pi}^t(q_l) > \Pi^{t+1}(q_l)$. For $t \geq m$, $\Pi^t(q_l) = 0$ which concludes the proof. □

Proof of lemma 5: It follows from lemma 4 that $A^t(q_l)$ is increasing in t. This implies that the optimal cutoff $k^t(q_l)$ defined in lemma 3 is decreasing in t. □

Proof of proposition 3: Note that $\tilde{\Pi}^0(0) > 0$. It will be shown that the derivative of $\tilde{\Pi}^t$ is less than δ for any $t \in \{0, \dots, m-1\}$. Since $\delta < 1$, this implies that $\tilde{\Pi}^0$ has a unique fixed point.

First, take $\tilde{\Pi}^{m-1}$. Since $\Pi^m = 0$, the expression in (10) simplifies. To show that the slope is less than δ, it is sufficient to show that $\tilde{\Pi}^{m-1}(\eta\Pi^0) - \tilde{\Pi}^{m-1}(\Pi^0) \leq \delta(\eta - 1)\Pi^0$ for any $\eta > 1$. This is equivalent to

$$k^m_\eta - k^m + \delta\eta\Pi^0\left((n - k^m_\eta)p_1 + \sum_{j=1}^{k^m_\eta} p_j\right) - \delta\Pi^0\left((n - k^m)p_1 + \sum_{j=1}^{k^m} p_j\right) \leq \delta n(\eta - 1)\Pi^0$$

where k^m_η is the optimal cutoff for $\eta\Pi^0$. Note that by lemma 3 $k^m_\eta \leq k^m$ and also note that the inequality is obviously true if $k^m_\eta = k^m$. For $k^m_\eta < k^m$, the previous expression can be rewritten as

$$\left(\sum_{k^m_\eta+1}^{k^m} p_1 - p_j - \frac{1}{\delta\Pi^0}\right) + (\eta - 1)\left(p_1(n - k^m_\eta) + \sum_{j=1}^{k^m_\eta} p_j\right) < n(\eta - 1).$$

This inequality holds true since by lemma 3 we have $p_1 - p_j \leq 1/(\delta\Pi^0)$ for all $j \leq k_m$.

Now we proceed by backwards induction. Assume that $\tilde{\Pi}^t(\Pi^0)$ is increasing in Π^0 with slope less than δ. Consider a reduction of Π^0 to $\eta\Pi^0$ with $\eta < 1$. If $k^t_\eta = k^t$, (10) implies directly that the slope of $\tilde{\Pi}^t$ is less than δ, i.e. that the reduction in $\tilde{\Pi}^t$ is less than $\delta(1-\eta)\Pi^0$. If, however, E can adjust k^t_η optimally, the loss from the reduction has to be even less. Hence, the slope of $\tilde{\Pi}^t$ is less than δ.

C's payoff when using the m-strategy and E best responds exceeds $1/(n-n\delta)$ strictly if $k^m < n$. (DERIVE CONDITION FOR THIS?) Otherwise, C's payoff in the m-equilibrium equals $1/(n-n\delta)$, i.e. the m equilibrium only exists if $V_O = 1/(n-n\delta)$. □

Proof of proposition 4: The proof is in a number of step. The following lemma which is also of independent interest is important for several of these steps.

Lemma 6. $\Pi^t(q_l)$ *is increasing in* q_l *in any* m*-equilibrium.*

Proof. Let $q_l' > q_l$. Denote by κ E's strategy that achieved the value $\Pi^t(q_l)$ after t consecutive misses and given quality q_l. This strategy consists of cutoff levels $\tilde{k}^t_{l+r}$ for all $t \in \{0,\dots,m-1\}$ and $r \geq 0$. Now suppose quality is q_l' and E used the strategy κ' after t consecutive misses where κ' is such that $k^t_{l'+r} = \tilde{k}^t_{l+r}$. Clearly, E's expected profits would then be larger than $\Pi^t(q_l)$ because he receives a more precise signal. This implies $\Pi^t(q_l') > \Pi^t(q_l)$. □

First, the posterior $(p_1(q_l),\dots,p_n(q_l))$ converges for l going to infinity. This is true as the cumulative distribution function increases in l while being bounded from above by the constant function 1. Hence, the cumulative distribution function converges and therefore each $p_k(q_l)$ converges. Using the limit posterior $(p_1^*,\dots,p_n^*)$ one can construct an m-equilibrium without learning as in section 4.1. Call the resulting equilibrium value of E after t misses Π^{t*}.

Second, $\Pi^t(q_l)$ converges as l grows to infinity (with $t \in \{1,\dots,m\}$). This follows from (i) the fact that $\Pi^t(q_l)$ is increasing in q_l and q_l is increasing in l and (ii) the fact that $\Pi^t(q_l)$ is bounded from a bove by $1/(1-\delta)$.

Third, $\lim_{l\to\infty} \Pi^t(q_l) = \Pi^{t*}$. As $\Pi^t(q_l)$ is increasing in quality, $\Pi^t(q_l) \leq \Pi^{t*}$ for all l and therefore also in the limit. It remains to show that $\lim_{l\to\infty} \Pi^t(q_l) < \Pi^{t*}$ is not possible. Denote the optimal cutoffs in the limit game (without learning) by k^{t*}. Consider the strategy of using the optimal cutoffs from the limit game $(k^{1*},\dots,k^{n*})$ in the game with learning (i.e. use cutoff k^{t*} after t consecutive misses regardless of the quality). Denote the expected profit stream of E when using this strategy as $\hat{\Pi}^t(q_l)$ (this is the expected discounted bonus stream after t consecutive misses when having quality q_l). As – for given cutoffs – E's profits are continuous in $p_k(q_l)$, it follows that $\lim_{l\to\infty} \hat{\Pi}^t(q_l) = \Pi^{t*}$. Given that $\hat{\Pi}^t(q_l) \leq \Pi^t(q_l)$, $\lim_{l\to\infty} \Pi^t(q_l) < \Pi^{t*}$ is impossible.

Fourth, L can be chosen such that $|\Pi^t(q_L) - \Pi^{t*}| < \varepsilon$ for all $t \in \{1,\dots,m\}$ by the third step. As $\Pi^t(q_l)$ is increasing in q_l and as $\Pi^{t*} \geq \Pi^t(q_l)$, this immediately implies $|\Pi^t(q_l) - \Pi^{t*}| < \varepsilon$ for all $l > L$. To conclude the same for $l > L$, recall that the slope of

$\tilde{\Pi}^{t-1}(\Pi^t)$ as defined by (10) is less than $\delta < 1$ (see proof of proposition 3). Consequently, deriving $\Pi^t(q_l)$ via backwards induction from $\Pi^t(q_L) = \Pi^{t*}$ will result in smaller than ε errors for $\Pi^t(q_l)$. Put differently, $|\Pi^t(q_L) - \Pi^{t*}| < \varepsilon$ implies that $|\Pi^t_L(q_l) - \Pi^{t*}| < \varepsilon$ for all $l < L$ where $\Pi^t_L(q_l)$ is the value derived via backwards induction from $\Pi^t(q_L) = \Pi^{t*}$. $\square$

Proof proposition 5: The result on $\Pi^t(q_l)$ is proven in lemma 6 above. For consumer surplus, note that $V^{m-1}(q_l)$ is strictly increasing in q_l under SM1: As $\Pi^{m-1}(q_l)$ is strictly increasing in q_l, $A^{m-1}(q_l)$ is increasing in q_l. Also $p_1(q_l) - p_k(q_l)$ is increasing in q_l by SM1. Hence, $k^{m-1}(q)$ is decreasing in q by lemma 3.

Consequently, C gets better information in $m-1$ if q_l is higher for two reasons: First, $k^{m-1}(q_l)$ could be lower. Second, even if the threshold stays the same, the probability that the option fitting C's needs is among the first k^{m-1} is higher for higher q_l because E's signal is more precise in the sense of first order stochastic dominance.

As shown in the proof of proposition 3, $\tilde{\Pi}^t(\Pi^0)$ has a slope of less than δ. This implies that $\Pi^t(q_l)$ increases in $\Pi^0(q_{l+1})$ with slope less than $\delta < 1$. Consequently, $\Pi^0(q_{l+1}) - \Pi^t(q_l)$ is increasing in l and therefore $A^t(q_l)$ is increasing in l. The same argument as above for $m-1$ shows then that $k^t(q_l)$ is decreasing and that C gets better information and therefore a higher expected utility after t consecutive misses if q_l is higher. Put differently, $V^t(q_l)$ increases in q_l. $\square$

References

Benabou, R. and G. Laroque (1992). Using privileged information to manipulate markets: Insiders, gurus, and credibility. *Quarterly Journal of Economics 107*(3), 921–958.

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica 50*(6), 1431–1451.

Park, I.-U. (2005). Cheap talk referrals of differentiated experts in repeated relationship. *RAND Journal of Economics 36*(2), 391–411.

Sobel, J. (1985). A theory of credibility. *Review of Economic Studies 52*(4), 557–573.