Competing for Boundedly Rational Consumers

Karl H. Schlag[1]

Febuary 2004

[1]The author wishes to thank Massimo Motta and Helder Vasconcelos for their comments. Additionally to the usual disclaimer the author wishes to point out that this version is not to be cited it is only a preliminary version for submitting to conferences.

Abstract

We consider a homogeneous good duopoly with rational firms facing consumers who minimize maximum regret. In each of a sequence of rounds, consumers have unit demand, may only visit a single firm and only observe the price charged by this firm where prices remain constant over time.

We illustrate the wide range of equilibrium prices supportable by simple consumer behavioral rules based on a reservation price strategy. Refining minimax regret to ensure time consistent behavior selects marginal cost pricing when consumers are moderately myopic.

In a further section we show how firms can achieve higher payoffs by setting prices stochastically.

JEL Classification: D43, L13, C72.

Keywords: regret, price competition, Bertrand paradox.

1 Introduction

As we find more interest in bounded rationality and as the methodologies for analyzing abstract decision- or game-theoretic settings have improved, it becomes natural to investigate how bounded rationality alters rational predictions in more realistic economic models. For the application we need to ask who is boundedly rational and to what degree and we need to select how bounded rationality will be modelled. Bounded rationality as defined as behavior that is not rational comes in many faces. It is selected or defined in many different ways: among others by fitting experimental data, by using axioms, by limiting strategy spaces in the rational context or by arguing plausibility. In this paper we do not want to give up the optimization paradigm when we leave the arena of rationality. We maintain the search for equilibria except that boundedly rational agents will follow a different objective than a rational one would (by definition). Models of bounded rationality also differ according to their degree of non rationality. We model inexperienced consumers as boundedly rational while the more experienced firms that have more to lose will be left rational. It is not uncommon to expect bounded rationality only when stakes are low and learning possibilities are insufficient.

There are other models in the literature that also consider rational firms and boundedly rational consumers. For instance, Rubinstein (1993) considers consumers with a very weak form of bounded rationality that have limited memory of previously observed prices. He shows how this can be utilized in equilibrium by a monopolist to price discriminate between consumers with different abilities. At the other end, Chintagunta and Rao (1996) and Hopkins (2003) consider consumers using exogenously specified rules and only look at firm optimization.

We consider the following very simple model of repeated competition in a homogeneous good oligopoly (in our analysis we limit attention to duopolies and triopolies). Firms are identical and compete in prices. Marginal costs are constant and strictly positive. Consumers are identical and in each round have unit demand with willingness to pay equal to one. Each consumer can only visit a single firm in each round, only observes prices of firms he visits and cannot store the good. Firms are infinitely patient and are restricted to setting the same price in each round. As there are no costs of visiting a different firm in the next round, any price between marginal cost and the monopoly price can be sustained in a subgame perfect equilibrium. We investigate which prices are sustainable in an equilibrium in which firms remain rational while consumers minimize

maximum regret for a given discount factor and are hence boundedly rational.

Regret is the difference between the payoffs you could get under perfect information and the payoffs that you actually get. *Minimax regret* was first introduced by Wald (1950) and captures the idea of learning when learning matters in a distribution free approach - there are no priors. The best outcome for a consumer under perfect information is to purchase forever at the firm offering the lowest price. What the consumer actually gets is determined by the given prices in the market and by the sequence of firms and purchasing decisions the consumer makes where payoffs are discounted over time. A consumer attains minimax regret if she chooses the behavior that minimizes over all behavioral rules her maximum regret over all possible price combinations. Minimax regret has proven a useful tool for determining behavior in decisions without specifying a prior (see Berry and Fristedt, 1985 and Schlag, 2003 among others) and for determining strategies and in games without specifying beliefs about opponent behavior (see Linhart and Radner, 1989).[1]

Since firms are restricted to charge the same price in each round it is best for any given consumer to purchase the good if its price is below one. So for consumers we only have to determine in equibrium which firm to visit in the next round given previous experience. Notice that consumer payoffs in each round are contained in the interval $[0,1]$.

First we consider a duopoly and present three simple rules that consumers could use, all based on a cutoff or reservation price strategy. Each rule prescribes to visit each firm equally likely in the first round. The simplest rule called the *simple cutoff rule* utilizes a single round memory and prescribes to go to the same firm again if its price is below a given threshold and to visit the other firm otherwise regardless of which firms have been visited in the past. The *grim trigger rule* is the standard trigger strategy that relies on slightly more memory in which the "trigger" is pulled if the price is above a given threshold. We find that the grim trigger rule yields the same expected payoffs as the simple cutoff rule. The *censored sampling rule* behaves like the simple cutoff rule whenever only one firm has been visited in the past but then visits forever the cheaper of the two firms once both firms have been visited.[2]

We characterize all symmetric equilibria supported by one of these three rules. There are two types of equilibria. In the first consumers shop in equilibrium forever at the same

[1] Maxmin has no predictive power for consumer behavior in this setting. The lowest payoff under any rule is realized when both firms charge the monopoly price. Thus any rule maximizes the minimumm payoff.

[2] The censored sampling rule can be used to support any price in $[c,1]$ as a subgame perfect equilibrium.

firm. Regret that the firm not visited might sell the product at price 0 results in an upper bound $1-\delta$ on the equilibrium price. There is also a strictly positive lower bound on the set of sustainable equilibrium prices driven by reluctancy to switch to the other firm when the price offered is low but above the equilibrium value. In the second type of equilibrium, consumers shop at both firms on the equilibrium path. While under the first two rules this yields monopoly pricing, marginal cost pricing results under the censored sampling rule as only the latter prescribes to only purchase at the cheapest firm. So we see that bounded rationality modelled by minimax regret restricts the set of equilibrium prices. The second effect of bounded rationality is that equilibria need not exist. In order for one of the above rules to attain minimax regret the discount value of the consumers cannot be not too large where the bounds are 0.5, 0.5 and 0.62 respectively. A fourth class of more sophisticated rules called *censored reinforcement rules* is added. These rules prescribe stochastic behavior after having visited only one firm that charges a high price. Equilibrium properties similar to those of the censored sampling rule are displayed except that the upper bound on the maximal discount value now equals 0.83.

Not to shop at the cheapest firm when both prices are known is not only nonintuitive, it also violates minimax regret conditional on this information. Notice that the original concept of minimax regret is ex-ante so the value of what is achievable is not updated when the consumer obtains more information. We refine minimax regret similar to subgame perfection to accommodate for time consistency which leads to the definition of *sequential minimax regret*. We find that there is essentially a unique rule that attains sequential minimax when $\delta \leq \frac{1}{2}$. This rule is a particular censored reinforcement rule. With this refinement only marginal cost pricing can be supported in a symmetric equilibrium. We also prove that no rule can attain sequential minimax regret when $\delta > \frac{1}{2}$.

Marginal cost pricing arises under sequential minimax regret due to the valuable information a consumer receives when he first visits a firm. The price charged by a firm once will always be charged by this firm by assumption. In the rest of the paper we investigate what happens if the firm has the possibility to prevent this inference by charging random prices. Formally we now allow for firms to choose mixed strategies in the game above. A mixed strategy is a price distribution according to which the price is independently drawn in each round. The most important impact is not on the equilibrium behavior of firms themselves but on consumers who now formulate maximum regret over all price distributions.

In this new setting we reconsider the above rules for consumer behavior and find

that none of them can attain minimax regret above a discount value of 0.38. To gain existence of equilibria for higher discount values we borrow two alternative rules from (Schlag, 2003). Under these rules we now also allow for three firms competing. Under the *simple reinforcement rule* the consumer visits each firm equally likely in the first round. Thereafter she visits the same firm again with probability equal to the payoff she obtained in the previous round. Otherwise she visits the next firm where the order of visits is chosen randomly before visiting the first firm. Under the *two state confidence rule* the consumer attributes two states associated to low and high confidence to each firm. The term confidence is used as states can be interpreted as beliefs a consumer could have on whether she has found a firm that offers a low expected price. According to this rule a consumer visits each firm equally likely in the first round and enters the low confidence state for the firm she visits. Conditional on being in the low state, with probability equal to the payoff obtained she visits the same firm again and enters the high state. Otherwise she visits a different firm and enters its low state. Conditional on being in the high confidence state she always visits the same firm again and transits between states as follows. With probability equal to her payoff she remains in the high state while she drops back down to the low state otherwise.

These two rules were discovered by Schlag (2003) as rules for $n = 2$ that maximize the upper bound on the discount value required for attaining minimax regret among rules with a single round and with two rounds of memory. The upper bound values are 0.41 and 0.62 respectively. In the appendix of this paper we derive the upper bounds for each of these rules when $n = 3$ and obtain the upper bounds 0.49 and 0.62 respectively. Applying these results to our particular application we find the following. Under the simple reinforcement rule only monopoly pricing is sustainable in equilibrium ($n = 2, 3$). The same result holds for the two state confidence rule if $n = 2$ or if $n = 3$ and $c \geq 1/2$. However, if $n = 3$ and $c < \frac{1}{2}$ then we obtain $2c$ as the unique expected price in a symmetric equilibrium. Three firms, low cost and two rounds of memory are sufficient to induce some competition between the firms. There is no restraint on how the precise form of the price distribution as long as its support is contained in $[0, 1]$ and its expected value equals $2c$. In particular, firms could be charging deterministic prices equal to $2c$ and they could be sometimes charging prices below marginal cost.

The paper is organized as follows. The main body starts with the setting of deterministic prices and presents the model, equilibrium definitions and separate subsections for the presentation and analysis of each rule. Then the framework with random prices

is introduced and previous rules are reconsidered before the two new rules from (Schlag, 2003) are presented. The main body ends with a brief conclusion. All calculations and proofs regarding minimax regret properties are contained in the appendix.

2 The Market

Consider n identical firms repeatedly selling a homogenous good to identical consumers. Firms have no fixed cost and constant marginal cost c with $0 < c < 1$. At the outset each firm i has to commit to charging the same (non negative) price in each round. Let p_i be the price charged by firm i. Firms aim to maximize long run average payoffs.

There is a finite number of identical consumers. Our results will not depend on the specific number of consumers. In each round, each consumer demands one unit of this good for which he is willing to pay at most a price equal to 1. Consumers do not know a priori any of the prices set by the individual firms. In a given round a consumer can only visit a single firm and consumers only observe the price of the firm they visit. A consumer gains utility $1 - p$ if she purchases the good at price $p \leq 1$ and gains utility 0 if she does not buy the good in that round. Behavior of a consumer is determined by a *purchasing rule* and a *shopping rule*. Throughout we consider the purchasing rule under which a consumer buys the good at price p if and only if $p \leq 1$. The shopping rule determines which firm the consumer visits in the next given his previous experience. As a consumer does not purchase the good above price 1 the payoffs of a consumer will be contained in $[0, 1]$.

We point out now (but leave the details until later) that any price in $[c, 1]$ can be sustained in a subgame perfect equilibrium in the traditional setting in which consumers are rational. However, in this paper consumers are boundedly rational and choose a behavior that attains minimax regret given the possible market prices using a given discount factor $\delta \in (0, 1)$. For more on minimax regret see the appendix. An *equilibrium* is a defined as a price p_i for each firm i and a shopping rule for each consumer such that consumers use a rule that minimizes maximum regret and each firm i has no incentive to unilaterally deviate from the specified price p_i. We will be only interested in symmetric equilibria in which all firms charge the same price and all consumers use the same rule.

3 Equilibrium Analysis

To simplify analysis we consider in this section only a duopoly (so $n = 2$). More importantly we refrain from a necessarily very involved general analysis of possible consumer in equilibrium. We will select a set of rules we informally argue as being simple in terms of amount of memory involved for the execution and in terms of how observed prices.

3.1 Deterministic Consumer Behavior

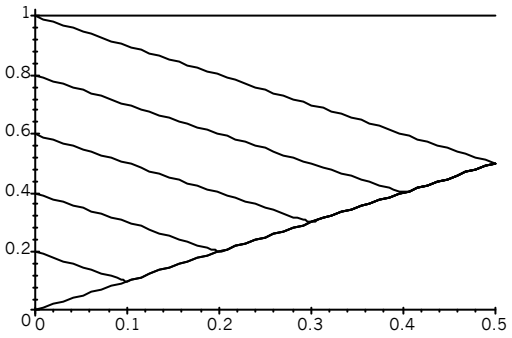
We start out by analyzing three rules or classes of rules. The three have in common that their choice in the first round is symmetric in the sense that each firm is visited equally likely and that they prescribe deterministic behavior thereafter (short: are *deterministic after round one*). The first rule builds on a very simple cutoff or reservation price strategy based only on information gathered in the previous round. The second is the standard grim trigger strategy. The third rule extends the first to incorporate memory of visits at both firms whenever available. All results on minimax regret behavior of the rules are stated and proven in the appendix.

We call the following behavior a *simple cutoff rule (with cutoff price α).* Visit each firm equally likely in the first round and in any later round visit the same firm again if and only if the price charged by this firm is at most α; $\alpha \in [0, 1]$ is exogenously given. Notice that in the appendix everything is defined in terms of payoffs so this rule is described in terms of a cutoff level y so $\alpha = 1 - y$.

Let us leave the setting of this paper for a moment to consider the more classic setting where all consumers are rational and where this is common knowledge. Then the simple cutoff rule can be used to support any price p_0 in $[c, 1]$ for any $\delta \in (0, 1]$. A symmetric Nash equilibrium is given if both firms charge the same price p and all consumers (who are rational) use the simple cutoff rule with cutoff price $\alpha = p$. One might argue that the simple cutoff rule is the simplest rule that has this property for any given price p in $[c, 1]$. Notice however that this Nash equilibrium is not subgame perfect.

Now return to the setting of this paper where consumers are not rational and instead aim to minimize maximum regret. We claim that both firms charging the same price p and all consumers using the simple cutoff rule with cutoff price $\alpha = p$ is an equilibrium if $\max\{c, \delta\} \leq p \leq 1 - \delta$ and hence $\delta \leq 1/2$. On the equilibrium path we find that each consumer forever visits only one firm. Firms cannot attract more consumers by reducing

their price. Raising own price also has negative consequences as all consumers then shop at the other firm. The fact that consumers attain minimax regret follows from our analysis in the appendix where we show that consumers attain minimax regret if $\delta \leq \alpha \leq 1 - \delta$. If $\delta \leq 1/2$ then it is also an equilibrium when all firms charge the monopoly price 1 and all consumers use the simple cutoff rule with a cutoff price $\alpha \in \left[\delta, \frac{1}{2}\right]$. In this equilibrium consumers constantly switch back and forth between the two firms. But each firm does not have an incentive to charge a different price p' that would make any consumer repeat her purchase as $p' \leq \frac{1}{2}$ would be necessary. It follows easily that no symmetric equilibria can be sustained when consumers use a simple cutoff rule provided that $\delta \leq 1/2$. As the simple cutoff rule does not attain minimax regret under deterministic prices for any other cutoff and discount values there are no other symmetric equilibria involving consumers using this rule.



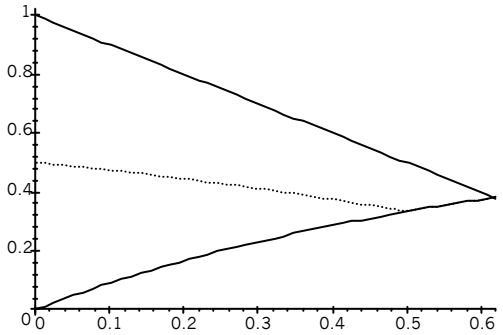
Set of prices sustainable in a symmetric equilibrium with a simple cutoff rule when $c \leq p$.

The second rule (or class of rules) we consider is called *grim trigger rule (with cutoff price α)*. As above this rule prescribes to visit each firm equally likely in the first round and to shop again at the same firm if the price charged by this firm is at most α. However, unlike the simple cutoff rule, should the price lie above α then the consumer shops forever at the other firm regardless of what price this firm charges. While this rule requires more memory it turns out that it generates the same expected payoffs as the simple cutoff rule with the same cutoff price α. Consequently, the same characterization of symmetric equilibria holds as obtained for the simple cutoff rule. The only difference concerns equilibrium consumer behavior under monopoly pricing where now the consumer

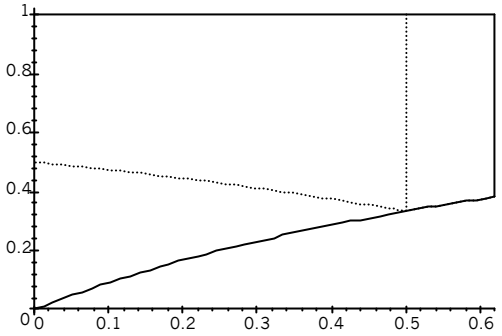
forever shops at the second firm visited while above the consumer constantly switched back and forth between the firms.

Our third rule is a bit more sophisticated and is called the *censored sampling rule (with cutoff price α)*. It prescribes the following behavior. Visit each firm equally likely in first round. Then visit the same firm again as long as the price in the previous round was at most α, otherwise visit the other firm in next round. When each firm has been visited at least once then visit in the next round the firm that charged the lowest price when visited last (visit same again if both charged the same price on the last visit). Notice for the classic setting in which all consumers are rational that this rule can sustain any price in $[c, 1]$ as a subgame perfect equilibrium.

Analogous arguments to the ones above now using Proposition 5 in the appendix show that it is an equilibrium if all firms charge price p and if all consumers use the censored sampling rule with cutoff price $\alpha = p$ provided $\max\left\{c, \frac{\delta}{1+\delta}\right\} \leq p \leq 1 - \delta$ and hence $\delta \leq \frac{1}{2}\sqrt{5} - \frac{1}{2}$ (where $\frac{1}{2}\sqrt{5} - \frac{1}{2} \approx 0.62$). The upper bound $1 - \delta$ and lower bound $\frac{\delta}{1+\delta}$ for the equilibrium price as functions of δ when c is sufficiently small are shown in the figure below. It is also an equilibrium if all firms charge price equal to marginal cost c and if all consumers use the censored sampling rule with cutoff price α provided that $\left\{\frac{\delta}{1+\delta}\right\} \leq \alpha \leq \min\{1 - \delta, c\}$ which means again that $\delta \leq \frac{1}{2}\sqrt{5} - \frac{1}{2}$. Failure of a censored sampling rule to attain minimax regret for alternative cutoff prices and discount values implies that no other symmetric equilibrium prices can be sustained with a censored sampling rule for given δ and c. In particular, and in contrast to the simple cutoff rule, the monopoly price cannot be supported with the censored sampling rule.



Prices strictly above marginal cost sustainable in a symmetric equilibrium under censored sampling lie between two solid lines. Adding sequential minimax regret predicts prices on dotted line when strictly above c.



Marginal cost pricing sustainable in a symmetric equilibrium under censored sampling lie within solid lines. Respective region using sequential minimax regret lie to the upper left of dotted lines.

In the minimax regret approach expected regret is calculated ex-ante, i.e. before visiting the first firm. Consumers need not minimize maximum regret conditional on having visited some firms and hence knowing some prices. For instance, once a consumer has visited both firms then regret from the perspective of that round is minimized if and only if the consumer only shops at a cheapest firm from then on. Yet neither the simple cutoff nor the grim trigger rule have this property. We say that a rule attains *sequential minimax regret* if maximum regret is minimized after any history similar to the way subgame perfection refines Nash equilibrium. The formal definition is provided in the appendix.

At this point of our analysis we use this refinement to select among the rules that are deterministic after round one (such as the three rules defined above). In the appendix we show that a rule that attains sequential minimax regret among the rules that are deterministic after round one exists if and only if $\delta \leq 1/2$ and that the censored sampling rule with cutoff price $\frac{1-\delta}{2-\delta}$ is essentially the unique rule with this property.[3] This refinement of consumer behavior hence selects the symmetric equilibrium with equilibrium price $p = \max\left\{c, \frac{1-\delta}{2-\delta}\right\}$ when $\delta \leq 1/2$ and yields no equilibrium when $\delta > \frac{1}{2}$. For illustration we include the equilibrium price $\frac{1-\delta}{2-\delta}$ attained for sufficiently small cost c in the figure above as a dotted line.

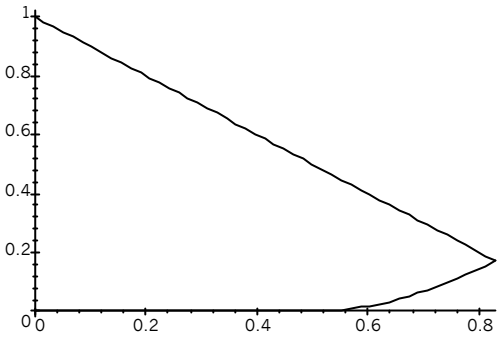
[3]The only degrees of freedom are how to behave when observing exactly the cutoff price and how to behave when both firms are observed to have set the same price.

3.2 Stochastic Consumer Behavior

In this section we extend our analysis to rules that are no longer deterministic after round one. First we consider a specific rule called the *censored reinforcement rule with cutoff price* α that prescribes the following behavior. Choose each firm equally likely in the first round. Whenever the same firm has been visited in each previous round then visit the same firm again if the price p is at most α and visit the same firm again with probability equal to $\frac{1-p}{1-\alpha}$ if $p > \alpha$, otherwise visit the other firm. Once each firm has been visited at least once then shop as prescribed by a censored sampling rule. Ignoring integer constraints we can also interpret the probability $\frac{1-p}{1-\alpha}$ of visiting the same firm again when $p > \alpha$ as visiting the firm visited in round one also in the next $\frac{1-p}{1-\alpha}\frac{1}{1-\frac{1-p}{1-\alpha}} = \frac{1-p}{p-\alpha}$ rounds and then visiting the other firm.

Equilibria with prices strictly above marginal costs can only be sustained if the equilibrium price equals the cutoff price. Otherwise the consumer eventually has observed both prices and shops forever at the cheapest firm. As firms are infinitely patient this then induces marginal cost pricing. Building on Proposition 6 in the appendix and using the same arguments as for the previous rules the following result is easily obtained. If $c \leq p \leq 1-\delta$ and either $\delta \leq \frac{1}{2}$ or $p \geq f(\delta) := \frac{4(1-\delta)+(2\delta-3)\sqrt{2(1-\delta)}}{\left(-2+\sqrt{2(1-\delta)}\right)\left(\delta-1+\sqrt{2(1-\delta)}\right)}$ and $\delta \leq 2\sqrt{2}-2$ (where $2\sqrt{2}-2 \approx 0.83$) then an equilibrium is given by all firms charging price p and all consumers using the censored reinforcement rule with cutoff price p. Competitive pricing can be supported in equilibrium by setting cutoff price α equal to 0 if $\delta \leq 1/2$ and by setting $\alpha = f(\delta)$ if $f(\delta) \leq c$ and $1/2 < \delta \leq 2\sqrt{2}-2$. No other symmetric equilibria exist when $\delta \leq 2\sqrt{2}-2$. Our lack of knowledge of minimax regret behavior for $\delta > 2\sqrt{2}-2$ prevents us to make any statements about what happens for outside this range of discount factors.

Notice that the equilibria found are very similar to those found under the censored sampling rule except that (i) the maximal feasible discount factor is now larger (0.83 verses 0.62) and (ii) for given discount factor below 0.62 the set of equilibrium prices is now larger as the constraint on the cutoff level from below is weaker, in particular when $\delta \leq 1/2$ then there is no lower bound on the cutoff price.



Upper and lower bound on equilibrium price sustainable under censored reinforcement.

Consider now sequential minimax regret behavior without restricting attention to deterministic rules as we did above. In the appendix we show for $\delta \leq 1/2$ that the censored reinforcement rule with cutoff price $\alpha = 0$ is essentially the unique rule that attains sequential minimax regret. Behavior of this rule after seeing only the price of a single firm is reminiscent of reinforcement as the same firm is visited again with probability equal to own payoff in that round. With this rule marginal cost pricing is the only sustainable equilibrium price. Unfortunately, as shown in the appendix, no rule attains sequential minimax regret when $\delta > \frac{1}{2}$.

3.3 Summary

We present four classes of rules that can be argued to be increasingly complex in terms of assumptions on memory and abilities to randomize. All four rules can be used to sustain any equilibrium price in $[c, 1]$ for any δ in the traditional setting with rational consumers. The latter two are also able to support any such price with a subgame perfect equilibrium. In our approach with boundedly rational consumers the set of equilibrium prices sustainable are very much restricted. The simpler rules can only attain minimax regret if δ is not too large. None of the rules can sustain prices above marginal cost that lie above $1 - \delta$ when $\delta < 0.828$. This result is due to the following. All rules only support equilibria in which prices lie above marginal cost in which in equilibrium consumers purchase forever at the first firm they visit. In the situation in which one

firm charges 0 and the other price p then their regret equals $\frac{1}{2}p$. This regret may not be larger than the maximal regret attainable which equals $\frac{1}{2}(1-\delta)$ when $\delta \leq 0.828$. So $\frac{1}{2}p \leq \frac{1}{2}(1-\delta)$ so $p \leq 1-\delta$.

Refining minimax regret by a adding time consistency condition yields extremely competitive behavior as it selects only marginal cost pricing. The drawback of this approach is that equilibria do not exist when $\delta > \frac{1}{2}$.

An alternative method for selecting equilibria is to consider for each class of rules the cutoff level that attains minimax regret for the largest range of discount values. The motivation is that boundedly rational consumers have no preferences among rules that attain minimax regret for a given discount factor and can use such a rule in more situations. With this refinement we present the selected cutoff level, resulting equilibrium price p and range of discount values:

Simple cutoff and grim trigger rule: $\alpha = \frac{1}{2}$ and $p = 1$ for $\delta \leq \frac{1}{2}$
Simple cutoff and grim trigger rule: $\alpha = p = \frac{1}{2}$ for $c \leq \frac{1}{2}$ and $\delta \leq \frac{1}{2}$
Censored sampling rule: $\alpha = \frac{3}{2} - \frac{1}{2}\sqrt{5}$ and $p = \max\left\{c, \frac{3}{2} - \frac{1}{2}\sqrt{5}\right\}$ for $\delta \leq \frac{1}{2}\sqrt{5} - \frac{1}{2}$,
Censored reinforcement rule: $\alpha = 3 - 2\sqrt{2}$ and $p = \max\left\{c, 3 - 2\sqrt{2}\right\}$ for $\delta \leq 2\sqrt{2} - 2$.

4 Random Pricing

In this section we investigate what happens when firms choose noisy prices or at least if consumers "fear" that prices may be random. More formally, at the beginning each firm i independently chooses a price distribution P_i. In each round the price p_i charged by firm i is then drawn according to P_i. Consumers calculate maximal regret over the set of all stationary price distributions.

As above firms charge prices below 1 in equilibrium as consumers will otherwise not purchase the good. Contrary to above, possibly firms occasionally charge prices below marginal cost as short run negative profits are of no concern to them. Thus $P_i \in \Delta[0,1]$. Firms remain infinitely patient and hence only care about long run profits.

We point out in the appendix that rules that attain sequential regret do not exist in this scenario. As prices may be random no history gives sufficient information to rule out any decision problem. This is advantageous for firms as above we show that sequential regret under deterministic pricing yields marginal cost pricing.

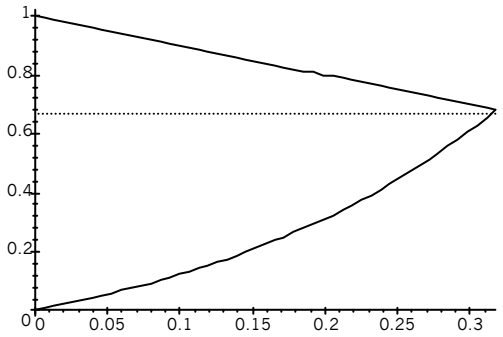
4.1 Previous Rules

Consider again the simple cutoff rule with cutoff level α. Proposition 3 in the appendix shows that it attains minimax regret (under random prices) if and only if $\delta \leq 1-\alpha \leq 1/3$. Thus all firms charging the same price p (with probability 1) and all consumers use the simple cutoff rule with cutoff level $\alpha = p$ is an equilibrium if $\max\{2/3, c\} \leq p \leq 1-\delta$ (so $\delta \leq \frac{1}{3}$). Monopoly pricing is sustainable if $\alpha = 2/3 < c$ and $\delta \leq 1/3$. Notice that the reservation price type of rule precludes any symmetric equilibrium to exist in which firms truly randomize.

$\delta \leq \frac{1}{3}$ is necessary in an equilibrium involving the simple cutoff rule as consumers calculate regret anticipating that prices could be random even though in equilibrium prices are set deterministically. Schlag (2003) shows that this upper bound on δ cannot be improved upon when considering only rules that rely only on events that occurred in the previous round.

Grim trigger allows for a wider range of cutoff prices. Following our results in the appendix, any price such that $\frac{\delta}{(1-\delta)^2} \leq p \leq 1-\delta$ (which implies $\delta < 0.318$) can be supported when consumers use the grim trigger rule with cutoff level $\alpha = p$.

In comparison to the simple cutoff rule, the upper bound on delta is slightly lower but the range of equilibrium prices has increased substantially. The figure below shows equilibrium prices sustainable for sufficiently small marginal cost c.



Upper and lower bound on sustainable equilibrium prices under sufficiently small marginal cost c for the grim trigger rule (solid line). Price lie above dotted line under the simple cutoff rule.

Under either censored sampling or censored reinforcement we find that $\delta \leq \frac{3}{2} - \frac{1}{2}\sqrt{5}$ (where $\frac{3}{2} - \frac{1}{2}\sqrt{5} \approx 0.38$) is necessary to attain minimax regret. Following our reasoning

above, price setting will be deterministic in any equilibrium supported by one of these two rules. We refrain from a necessarily involved complete characterization as the upper bound of 0.38 on the discount factor seems extremely restrictive and does not improve substantially on the upper bound on δ under the simple cutoff rule.

To summarize, we find that the previous rules that were designed to deal with deterministic prices can only be used for limited ranges of discount factors. In an environment with random prices consumers receive much less information when observing a single or even a finite number of prices charged by the same firm. It is then intuitive that their use of cutoff prices does not do well at aggregating information and is hence less adapted for learning when payoffs are random.

4.2 Linear Rules

We now present two rules that both exhibit random behavior after round one where randomizations (formally the probability of choosing an action) is linear in payoffs received. In this analysis we also allow for more than two firms.

4.2.1 Simple Reinforcement

Consider the following rule that combines minimal use of memory as demonstrated by the simple cutoff rule with payoff reinforcing behavior as exhibited by the censored reinforcement rule. The rule is called the *simple reinforcement rule* and prescribes the following behavior. Initially, randomly assign indices to the n firms. In the first round visit each firm equally likely. From round two on visit the same firm again with probability equal to $1-p$ where p is the price charged by this firm in the previous round. With probability p visit the firm with the next higher index (modulo n). Let q_i be the long run average visits at firm i. Then

$$\begin{aligned} q_i &= \frac{1}{\bar{p}_i \sum_{k=1}^{n} \frac{1}{\bar{p}_k}} \text{ if } \min_k \{\bar{p}_k\} > 0 \\ q_i &= \frac{1}{\#\{k : \bar{p}_k = 0\}} \text{ if } \bar{p}_i = 0 \\ q_i &= 0 \text{ if } \bar{p}_i > 0 = \min_k \{\bar{p}_k\} \end{aligned}$$

where long run average payoffs of firm i equal $q_i(\bar{p}_i - c)$. When all expected prices are equal then each firm is visited equally likely. Firms that are cheaper on average are visited

more likely. The most expensive firm obtains customers even in the long run as long as no other firm charges a (deterministic) price equal to the lower end of the price interval.

Consider an equilibrium. Let R_i denote the long run average payoff of firm i in an equilibrium. As $\bar{p}_i \geq c > 0$ in equilibrium we find

$$R_1 = \frac{\bar{p}_1 - c}{1 + \bar{p}_1 \sum_{k=2}^{n} \frac{1}{\bar{p}_k}}$$

where

$$\frac{d}{d\bar{p}_1} R_1 = \frac{1 + c \sum_{k=2}^{n} \frac{1}{\bar{p}_k}}{\left(1 + \bar{p}_1 \sum_{k=2}^{n} \frac{1}{\bar{p}_k}\right)^2} .$$

Given $\frac{d}{d\bar{p}_1} R_1 > 0$ if $\bar{p}_1 < 1$ we obtain the following result.

Proposition 1 *There exists $\delta_n \in (0,1)$ where $\delta_2 = \sqrt{2} - 1 \approx 0.41$ and $\delta_3 = \frac{1}{2}\sqrt{4\sqrt{3} - 3} - \frac{1}{2} \approx 0.49$ such that if $\delta \leq \delta_n$ then firm i choosing a mixed price according to distribution $P_i \in \Delta[0,1]$ for each i and consumers using the simple reinforcement rule is an equilibrium if and only if each firm sets price 1, i.e. $P_i(1) = 1$ for all i.*

The above result is disappointing in two respects. (1) Equilibria do not exist for discount values above $\frac{1}{2}$ as the simple reinforcement rule has a restricted range of discount values under which it attains minimax when $n = 2, 3$. For $n = 2$ we know that this is due to the simple memory of this rule. (Schlag, 2003) shows that there is no rule with a single round memory that can attain minimax regret for $\delta > \delta_2$. (2) Only monopoly pricing arises when equilibria exist. There is too little learning keep ensure that firms actually compete. Notice that in equilibrium consumers alternate in each round between the two firms.

4.2.2 Two State Confidence

In the following we consider a rule that has two rounds of memory. This rule is the so-called *two state confidence rule*. For each firm that the consumer visits there are two states which can be interpreted as confidence levels low and high. High confidence intuitively refers to the fact that the consumer qualifies this firm as one that offers low prices on average. The rule prescribes the following behavior. As above, initially assign indices at random to the n firms. In the first round visit each firm equally and enter the low confidence state of the respective firm. In any later round, after previously being in the

low confidence state of a firm, with probability equal to the payoff $1-p$ received in the previous round visit the same firm again and transit to the high state. Otherwise, i.e. with probability p, visit the firm with the next higher index (modulo n) and enter its low confidence state. Thus you never are in the low confidence state of the same firm in two consecutive rounds. In any later round, after previously being in the high confidence state of a firm, visit the same firm again in the next round. Remain in the high state with probability $1-p$ and transit to the low confidence state of the same firm with probability p.

Given q_i denotes long run average number of visits to firm i we obtain

$$\begin{aligned} q_i &= \frac{1}{\bar{p}_i^2 \sum_{j=1}^{n} \frac{1}{\bar{p}_j^2}} \text{ if } \min_k \{\bar{p}_k\} > 0 \ , \\ q_i &= \frac{1}{\#\{k : \bar{p}_k = 0\}} \text{ if } \bar{p}_i = 0 \ , \\ q_i &= 0 \text{ if } \bar{p}_i > 0 = \min_k \{\bar{p}_k\} \ . \end{aligned}$$

Assume in equilibrium that $\bar{p}_i = \bar{p}$ for $i \geq 2$. Then

$$R_1 = \frac{1}{\bar{p}_1^2 \left(\frac{1}{\bar{p}_1^2} + \frac{n-1}{\bar{p}^2} \right)} (\bar{p}_1 - c) = \frac{\bar{p}^2}{\bar{p}^2 + (n-1)\bar{p}_1^2} (\bar{p}_1 - c)$$

and

$$\frac{d}{d\bar{p}_1} R_1 = \bar{p}^2 \frac{\bar{p}^2 + 2(n-1)\bar{p}_1 c - (n-1)\bar{p}_1^2}{\left(\bar{p}^2 + (n-1)\bar{p}_1^2\right)^2}$$

where $\frac{d}{d\bar{p}_1} R_1 \geq 0$ if $n = 2$ and where $\bar{p}^2 + 2(n-1)\bar{p}c - (n-1)\bar{p}^2 = 0$ holds if $\bar{p} = \frac{2(n-1)}{n-2} c$ and $n \geq 3$. As we again are only able to verify the minimax properties of the above rule for $n \leq 3$ we can only present a result for these values of n.[4]

Proposition 2 *Assume $n \in \{2, 3\}$ and $\delta \leq \frac{1}{2}\sqrt{5} - \frac{1}{2}$ (≈ 0.62). Firm i choosing a distribution $P \in \Delta[0,1]$ with expected price $\bar{p}$ for all i and consumers using the two state confidence rule is a symmetric equilibrium if and only if*

$$\bar{p} = \begin{cases} 1 \text{ if } n = 2 \text{ or } [n = 3 \text{ and } c \geq \frac{1}{4}] \\ 4c \text{ if } n = 3 \text{ and } c < \frac{1}{4} \end{cases} .$$

[4] It is easily argued that both the simple reinforcement rule and the two state confidence rule attain minimax regret if δ is sufficiently small. However, as we have no explicit proof of the analytical value of this bound such a result is of little added value.

5 Conclusion

This paper contains a simple model to investigate equilibria with different degrees of rationality on either side of the market. While consumers are boundedly rational, we want to emphasize that their behavior is not assumed exogenously. Instead they still solve an optimization problem albeit not the one associated to rational agents. Equilibria as in the approach of (Linhart and Radner, 1988) are investigated. We refine minimax regret to ensure time consistent behavior. This eliminates nonintuitive behavior as displayed by the simple cutoff rule and by the grim trigger rule arising when the consumer knows the prices of both firms. The refinement is however too strong when prices are random and future research may find other means to select behavior.

Rules found in (Schlag, 2003) attain minimax regret under deterministic prices. However these rules have been developed for the random payoff scenario. Others are more appealing when payoffs are deterministic. So we derive alternative rules that attain minimax regret under deterministic payoffs and put this material in the appendix. Despite the length of the paper, we find that this material should not be extracted as it is part of the optimization in the equilibrium.

Conditions to obtain minimax regret are intricate. We avoid presenting an even longer paper by only considering a duopoly for the case where prices are deterministic. More general results for the case of random prices are not available and under current understanding of this area of research are extremely difficult to obtain.

Despite the difficulty of finding rules that attain minimax regret we believe that the rules presented are plausible for consumers for two reasons. (i) They are natural candidates even if their nice properties were not known. (ii) They are applicable to very general settings and do not have to be reoptimized each time a new situation is faced. They apply to any decision problem in which payoffs are known to be contained in a bounded connected interval. This interval is then linearly rescaled to become $[0, 1]$. One might therefore argue informally that nice properties of a minimax regret rule are actually easier to discover than alternative properties of more specialized rules.

Our market model is very simple and easily criticized. One concern could be that prices set by firms are fixed forever. For the setting of the first part of the paper where random prices are not allowed this is not implausible. We can easily imagine external restrictions that keep a firm from changing its prices each day. Moreover, it is easily shown that a single unilateral deviation by some firm in a later round will not be profitable when

starting in an equilibrium. An interested reader will see that this holds for any of the rules presented. Of course for the setting that allows for random prices it is more natural to also investigate non stationary pricing which we plan on doing. Here we view our modelling at least as an important first step. A less persuasive argument in favor of stationary prices is that this is also the approach underlying a logit equilibrium (e.g. see Anderson et al. 2000).

The specific form of competition taking place in our market model was chosen to allow for use of existing results on minimax regret behavior. The more classic model with sequential search before making a single purchase has a different payoff structure. This is left for future research as it requires initially an extensive analysis of behavioral rules. In our market model the rational predictions of market prices are extremely weak. In contrast we find that simple rules put both upper and lower bounds on sustainable prices. There may be other rules that can support all prices in $[c, 1]$. We do not follow this line of research because we are interested in simple rules. Thus we were very careful in selecting the rules in this paper.

One of the most interesting predictions of our model is that fear of stochastic pricing combined with intermediate degree of myopia can make a consumer choose to only react weakly to changes in prices which in turn enables firms to charge higher prices. In our simple model these prices are in fact maximal unless there are three instead of two firms and costs are small.

References

[1] Anderson, S.P., J.K. Goeree and C.A. Holt (2000), "The logit equilibrium - A perspective on intuitive behavioral anomalies", forthcoming in *Southern Economic Journal*.

[2] Berry, D.A. and B. Fristedt (1985), *Bandit Problems: Sequential Allocation of Experiments*, Chapman-Hall, London.

[3] Chintagunta, P.K., and P.V.Rao (1996), "Pricing strategies in a dynamic duopoly: a differential game model," *Management Science* **42**, 1501-14.

[4] Hopkins, E. (2003) *Adaptive Learning Models of Consumer Behaviour*, Mimeo, University of Edinburgh.

[5] Kreps, D. and J. Scheinkman (1983) "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes," *Bell Journal of Economics* **14**, 326-337.

[6] Linhart, P.B. and R. Radner (1989) "Minimax-Regret Strategies for Bargaining over Several Variables," *J. Econ. Theory* **48**, 152-178.

[7] Rubinstein A. (1993) "On Price Recognition and Computational Complexity in a Monopolistic Model," *J. Pol. Econ.* **101**, 473-484.

[8] Schlag, K.H. (2003) "How to Choose, A Boundedly Rational Approach to Repeated Decision Making," Mimeo, European University Institute, http://www.iue.it/Personal/Schlag/papers/regret7.pdf.

[9] Wald, A. (1950), *Statistical decision functions*, Chelsea: Bronx.

A Choosing a Rule

A.1 Minimax Regret

In the following we show how to select behavior using minimax regret. $D = (P_i)_{i=1}^n$ is called a *decision problem* if P_i is a probability measures on $[0,1]$ for $i = 1,..,n$ where π_i denotes the expected payoff realized by P_i. D is deterministic when $P_i(x_i) = 1$ for some $x_i \in [0,1]$ for all i. Let $\Delta[0,1]^n$ denote the set of decision problems and let $[0,1]^n$ denote the subset of *deterministic decision problems* in which each action yields a deterministic payoff. Consider an individual repeatedly and independently facing the same decision problem. A *behavior rule* is a function $f : \cup_{k=0}^{\infty}(\{1,..,n\} \times [0,1])^k \to \Delta\{1,..,n\}$ that determines for any history of actions and payoffs received which action to choose in next round where $\Delta\{1,..,n\}$ is the set of probability distributions over the set of actions. Let $\mathcal{F}$ denote the set of behavioral rules. We say that a behavioral rule f is *deterministic after round* 1 if $f\left((\{1,..,n\} \times [0,1])^k\right) \to \{1,..,n\}$ when $k \geq 1$.

Assume that the individual discounts future payoffs with discount factor $\delta \in (0,1)$ so she evaluates the stream of payoffs $(x_k)_{k=1}^{\infty}$ by calculating $(1-\delta)\sum_{k=1}^{\infty}\delta^{k-1}x_k$. Any given rule f generates against any given decision problem D expected discounted payoffs

$$\pi(D,f) = (1-\delta)\sum_{k=1}^{\infty}\delta^{k-1}\sum_{i=1}^{n}p_i^{(k)}(D,f)\,\pi_k(D)$$

where $p_i^{(k)}(D,f)$ is the probability of choosing action i in round k. We assume that the individual aims to maximize expected payoffs if it knew which decision problem it is facing. Thus, if faced with D then she would choose in each round an action from $\arg\max_{i=1,..,n}\{\pi_i(D)\}$. However we assume that the individual does not know which decision problem from $\Delta[0,1]^n$ she is facing. Moreover we do not consider a prior over decision problems. Instead we assume that she chooses the rule that minimizes the maximum expected difference between what she could get and what she does get. Formally, we say that the behavioral rule f *attains minimax regret* (Wald, 1950, Berry and Fristedt, 1985) if $f \in \arg\min_{f\in\mathcal{F}} \sup_{D\in\Delta[0,1]^n} r(D,f)$ where $r(D,f) = \max_{i=1,..,n}\pi_i(D) - \pi(D,f)$ denotes the expected *regret* when choosing rule f against decision problem D. In parts of our analysis we only restrict attention to deterministic decision problems and then say that f *attains minimax regret under deterministic payoffs* if $f \in \arg\min_{f\in\mathcal{F}} \sup_{D\in[0,1]^n} r(D,f)$. We also sometimes restrict attention to a subset of rules such as those that are deterministic after round 1 and then add a subfix and say "attains minimax regret among the rules that are deterministic after round 1."

In the following sections we investigate minimax properties of specific rules. The first three rules are deterministic after round 1. Unless otherwise specified we limit attention to the case of two actions.

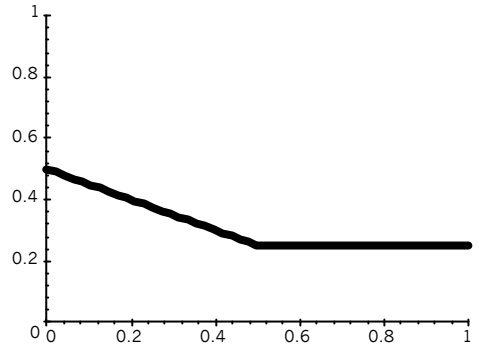
A.2 Simple Cutoff Rules

The first rule we call the *simple cutoff rule (with cutoff level y)* specifies to choose each action equally likely in the first round and from then on to choose the same action again if and only if the payoff received is at least y where $y \in [0,1]$ is exogenously given. This rule requires no memory from earlier rounds. If payoffs are deterministic then

$$\pi^\delta = \begin{cases} \frac{1}{2}(\pi_1+\pi_2) \text{ if } \min\{\pi_1,\pi_2\} \geq y \\ \frac{1}{2}\pi_1 + \frac{1}{2}((1-\delta)\pi_2 + \delta\pi_1) \text{ if } \pi_1 \geq y > \pi_2 \\ \frac{1}{2}(\pi_1+\pi_2) \text{ if } y > \max\{\pi_1,\pi_2\} \end{cases}.$$

So $\sup_{\pi_1,\pi_2\in[0,1]} r = \max\left\{\frac{1}{2}(1-y), \frac{1}{2}(1-\delta), \frac{1}{2}y\right\}$ where the three terms coincide to the three cases in the formula of π^δ above. Proposition 6 below shows that the value of minimax regret equals $\frac{1}{2}(1-\delta)$ for $\delta \leq 2\sqrt{2}-2 \approx 0.83$. Hence, given $\delta \leq 2\sqrt{2}-2$ the simple cutoff rule attains minimax regret under deterministic payoffs if and only if $\delta \leq y \leq 1-\delta$ so $\delta \leq 1/2$ where the highest value $\delta = 1/2$ is achieved when $y = 1/2$.

When $\delta > \frac{1}{2}$ then it follows immediately that the simple cutoff rule with cutoff level $\frac{1}{2}$ is the unique rule that attains minimax regret under deterministic payoffs among the simple cutoff rules. The maximal regret for this rule is shown in the following figure.



Comparing the maximal regret of this rule to that found in Proposition 6 shows that there is no simple cutoff rule that attains minimax regret when $\delta > \frac{1}{2}$.

Now assume instead that payoffs are random. It is can be argued (for more details see Schlag, 2003) that it is sufficient consider only decision problems in which $P_1(1) = q_1 = 1 - P_1(y - \varepsilon)$ and $P_2(y) = q_2 = 1 - P_2(0)$ which yields regret that in the limit as ε tends to 0 tends to

$$\tilde{r}(q_1, q_2) = \frac{1}{2} \frac{1 + \delta - 2\delta q_1}{1 + \delta - \delta(q_1 + q_2)} (q_1 + (1 - q_1) y - q_2 y) \ .$$

Schlag (2003, see also Proposition 10) shows that the value of minimax regret equals $\frac{1}{2}(1 - \delta)$ for all $\delta \leq \frac{1}{2}\sqrt{5} - \frac{1}{2}$. As $\tilde{r}(1, 0) = \frac{1}{2}(1 - \delta)$ we investigate when $(1, 0) \in \arg\max \tilde{r}$. A necessary condition is that $\frac{d}{dq_1}\tilde{r}(1, 0) \geq 0 \geq \frac{d}{dq_2}\tilde{r}(1, 0)$ which is easily shown to be equivalent to $\delta \leq y \leq 1/3$. Verifying for $\delta \leq y \leq 1/3$ that $\frac{d}{dq_2}\tilde{r} \leq 0$ and that $\frac{d}{dq_1}\tilde{r}(q_1, 0) \geq 0$ completes the proof of the following result.

Proposition 3 *Assume $n = 2$. The simple cutoff rule with cutoff level y (i) attains minimax regret under deterministic payoffs if and only if $\delta \leq y \leq 1 - \delta$ (so $\delta \leq \frac{1}{2}$) and (ii) attains minimax regret under random payoffs if $\delta \leq y \leq \frac{1}{3}$ but not for any other values of y and δ if $\delta \leq \frac{1}{2}\sqrt{5} - \frac{1}{2}$ (≈ 0.62).*

A.3 Grim Trigger Rules

Next consider the *grim trigger rule with cutoff* y that prescribes the following behavior. Choose each action equally likely in the first round. After having always chosen the same action, choose the same action again if the last payoff was at least as large as y. Choose the other action and never switch back if the first chosen action yielded a payoff below y. Note that this rule requires memory of whether a different action has been chosen in the past.

If payoffs are deterministic then we obtain the same expression for the expected payoffs π^δ as under the simple cutoff rule. Hence the necessary and sufficient conditions on y and δ for attaining minimax regret under deterministic payoffs are the same as those for the simple cutoff rule.

Now assume that payoffs are random. Let x_i be the payoff to choosing action i after choosing action i in all previous rounds. Let q_i be the probability that action i yields a payoff above y. Then $x_i = (1-\delta)\pi_i + q_i\delta x_i + (1-q_i)\delta\pi_{3-i}$. Let z_i be the payoff to choosing action i in round one then

$$\begin{aligned} z_i &= (1-\delta)\pi_i + q_i\delta x_i + (1-q_i)\delta\pi_{3-i} \\ &= (1-\delta)\pi_i + q_i\delta\frac{1}{1-\delta q_i}\left((1-\delta)\pi_i + (1-q_i)\delta\pi_{3-i}\right) + (1-q_i)\delta\pi_{3-i} \\ &= \frac{(1-\delta)\pi_i + (1-q_i)\delta\pi_{3-i}}{1-\delta q_i} \end{aligned}$$

As $\pi^\delta = \frac{1}{2}(z_1+z_2)$ and assuming $\pi_1 \geq \pi_2$ we obtain

$$\begin{aligned} r &= \pi_1 - \frac{1}{2}\left(\frac{(1-\delta)\pi_1 + (1-q_1)\delta\pi_2}{1-\delta q_1} + \frac{(1-\delta)\pi_2 + (1-q_2)\delta\pi_1}{1-\delta q_2}\right) \\ &= \frac{1}{2}(\pi_1-\pi_2)\frac{\left(1-2\delta q_1 + \delta^2(q_1-q_2) + \delta^2 q_1 q_2\right)}{(1-\delta q_1)(1-\delta q_2)} \end{aligned}$$

which attains its supremum at

$$\bar{r}(q_1,q_2) := \frac{1}{2}(q_1 + (1-q_1)y - q_2 y)\frac{\left(1-2\delta q_1 + \delta^2(q_1-q_2) + \delta^2 q_1 q_2\right)}{(1-\delta q_1)(1-\delta q_2)}.$$

We now need to find conditions under which $\bar{r}$ is maximized when $q_1 = 1$ and $q_2 = 0$. Note that $\frac{d}{dq_1}\bar{r}(1,0) = \frac{1}{2}\frac{1-y+(2y-3)\delta+(1-y)\delta^2}{1-\delta} \geq 0$ if and only if $y \leq \frac{1-3\delta+\delta^2}{(1-\delta)^2}$ and $\frac{d}{dq_2}\bar{r}(1,0) = -\frac{1}{2}(1-\delta)(y-\delta) \leq 0$ if and only if $y \geq \delta$ where $\frac{1-3\delta+\delta^2}{(1-\delta)^2} \geq \delta$ implies $\delta \leq 1 - \sqrt[3]{\frac{1}{2} + \frac{1}{18}\sqrt{93}} +$

$\frac{1}{3\sqrt[3]{\frac{1}{2}+\frac{1}{18}\sqrt{93}}} \approx 0.3177$. Few additional steps show that $\frac{d}{dq_2}\bar{r} \leq 0$ and that $\frac{d}{dq_1}\bar{r}(q_1, 0) \geq 0$ holds when $y \leq \frac{1-3\delta+\delta^2}{(1-\delta)^2}$. We summarize.

Proposition 4 *Assume $n = 2$. (i) The grim trigger rule with cutoff level y attains minimax regret under deterministic payoffs if and only if $\delta \leq y \leq 1-\delta$ (so $\delta \leq \frac{1}{2}$). (ii) The grim trigger rule with cutoff level y attains minimax regret under random payoffs if $\delta \leq y \leq \frac{1-3\delta+\delta^2}{(1-\delta)^2}$ (so $\delta < 0.318$) but not for any other values y and δ if $\delta \leq \frac{1}{2}\sqrt{5}-\frac{1}{2}$.*

Notice that in order to attain minimax regret under random payoffs the grim trigger rule requires slightly lower discount values than the simple cutoff rule. However there are many more feasible cutoff levels for given δ under the grim trigger rule. This plays a role for our analysis of price setting behavior of firms.

A.4 Censored Sampling Rules

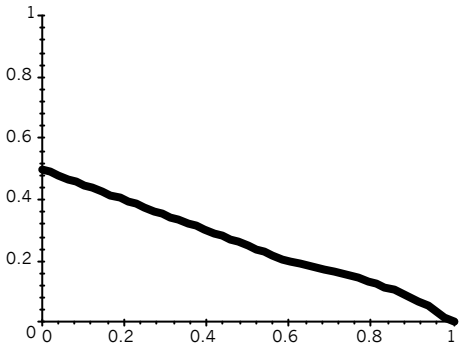
Next consider slightly more sophisticated behavior that generalizes the simple cutoff rule. This rule we call the *censored sampling rule with censor cutoff y*. Choose each action equally likely in the first round. Then choose same action again as long as payoff in previous round is at least y, otherwise choose other action in next round. Once each action has been chosen at least once, choose in next round the action that yielded the highest payoff when tried last (choose same again if both achieved same payoff). This rule requires memory of the last payoff obtained by each action previously chosen. When $y = 1$ and payoffs are deterministic and below 1 then first each action is chosen only once and then the action that attained a higher payoff is chosen forever.

Assume payoffs are deterministic. Then

$$\pi^\delta = \begin{cases} \frac{1}{2}(\pi_1+\pi_2) \text{ if } \min\{\pi_1,\pi_2\} \geq y \\ \frac{1}{2}\pi_1 + \frac{1}{2}((1-\delta)\pi_2 + \delta\pi_1) \text{ if } \pi_1 \geq y > \pi_2 \\ \frac{1}{2}\left(1-\delta^2\right)(\pi_1+\pi_2) + \delta^2 \max\{\pi_1,\pi_2\} \text{ if } y > \max\{\pi_1,\pi_2\} \end{cases}.$$

So $\sup_{\pi_1,\pi_2\in[0,1]} r = \max\left\{\frac{1}{2}(1-y), \frac{1}{2}(1-\delta), \frac{1}{2}\left(1-\delta^2\right)y\right\}$ where the three terms coincide to the three cases in the formula of π^δ above. As we know that minimax regret value equals $\frac{1}{2}(1-\delta)$ for $\delta \leq 2\sqrt{2}-2$ and that this rule chooses a best response when $\{\pi_1,\pi_2\} = \{0,1\}$ we obtain that this rule attains minimax regret under deterministic payoffs if $\delta \leq \frac{1}{2}\sqrt{5}-\frac{1}{2}$ and $\delta \leq y \leq \frac{1}{1+\delta}$.

If we limit our selection to censored sampling rules then $\frac{1}{2}(1-y) = \frac{1}{2}\left(1-\delta^2\right)y$ is solved by $y = \frac{1}{2-\delta^2}$. So if $\delta > \frac{1}{2}\sqrt{5} - \frac{1}{2}$ then the censored sampling rule with cutoff $\frac{1}{2-\delta^2}$ is the unique rule that attains minimax regret under deterministic payoffs among the censored sampling rules. Maximum regret of this rule equals $\frac{1}{2}\frac{1-\delta^2}{2-\delta^2}$. In particular, our results above show that setting $y = 1$, which seems a natural behavior for large δ, does not yield minimax regret under deterministic payoffs for any value of δ. The value of minimax regret among the censored sampling rules is shown in the following figure.



Consider now random payoffs. Consider a decision problem in which only payoffs in $\{0,1\}$ are realized. Then the censored sampling rule behaves like the grim trigger rule independent of any specification of cutoff levels. Entering $q_1 = \pi_1$ and $q_2 = \pi_2$ in the expression for r found in Section A.3 and then looking at $\frac{d}{d\pi_1} r(1,0)$ it is easily verified that the censored sampling rule does not attain minimax regret under random payoffs if $\frac{3}{2} - \frac{1}{2}\sqrt{5} < \delta \leq \frac{1}{2}\sqrt{5} - \frac{1}{2}$. We refrain from a complete analysis of this rule as it is intricate and will not add sufficiently to the results of this paper.

Proposition 5 *Assume $n = 2$. (i) The censored sampling rule with cutoff level y attains minimax regret under deterministic payoffs if and only if $\delta \leq y \leq \frac{1}{1+\delta}$ (so $\delta \leq \frac{1}{2}\sqrt{5} - \frac{1}{2}$). (ii) There is no censored sampling rule that attains minimax regret under random payoffs if $\frac{3}{2} - \frac{1}{2}\sqrt{5} < \delta \leq \frac{1}{2}\sqrt{5} - \frac{1}{2}$ (where $\frac{3}{2} - \frac{1}{2}\sqrt{5} \approx 0.38$).*

A.5 Censored Reinforcement Rules

Our next rule is even more sophisticated as unlike the previously defined rules it is no longer deterministic after round 1. Let us call this rule the *censored reinforcement rule*

with cutoff level y. Accordingly, choose each action equally likely in the first round. Choose the same action again if last payoff x is at least y or with probability x/y if $x < y$ and switch to other action with probability $1 - \frac{x}{y}$ if $x < y$. Notice that this rule requires the same memory requirements as the censored sampling rule. The special case where $y = 1$ has a nice interpretation and will play an important role in a later section. Under this rule, as long as the other action has not been chosen, one might say that payoffs reinforce choice of the same action as the same action is chosen again with probability equal to the payoff obtained.

Assuming deterministic payoffs it is easily verified that

$$\pi^{\delta} = \begin{cases} \frac{1}{2}(\pi_1 + \pi_2) \text{ if } \min\{\pi_1, \pi_2\} \geq y \\ \frac{1}{2}\pi_1 + \frac{1}{2}\frac{(1-\delta)\pi_2 + \left(1 - \frac{\pi_2}{y}\right)\delta\pi_1}{1 - \delta\frac{\pi_2}{y}} \text{ if } \pi_1 \geq y > \pi_2 \\ \pi_1 - \frac{1}{2}(1-\delta)(\pi_1 - \pi_2)\frac{(1+\delta)y^2 - 2\delta\pi_1 y + \pi_2(\pi_1 - y)\delta^2}{(y - \delta\pi_1)(y - \delta\pi_2)} \text{ if } y \geq \pi_1 \geq \pi_2 \end{cases}$$

If $\pi_1 \geq \pi_2 \geq y$ then $r = \frac{1}{2}(\pi_1 - \pi_2)$ which is maximally $\frac{1}{2}(1-y)$. If $\pi_1 \geq y > \pi_2$ then $r = \frac{1}{2}y(1-\delta)\frac{\pi_1 - \pi_2}{y - \pi_2\delta}$ with supremum equal to $\max\left\{\frac{1}{2}(1-\delta), \frac{1}{2}(1-y)\right\}$. Finally, if $y \geq \pi_1 \geq \pi_2$ then $r = r(\pi_1, \pi_2) = \frac{1}{2}(1-\delta)(\pi_1 - \pi_2)\frac{\delta^2\pi_1\pi_2 - 2\delta\pi_1 y - \pi_2 y\delta^2 + \delta y^2 + y^2}{(y - \delta\pi_1)(y - \pi_2\delta)}$ for which it is easily verified that $\frac{d}{d\pi_2}r(\pi_1, \pi_2) \leq 0$, that $r(\pi_1, 0)$ is concave and that $\frac{d}{d\pi_1}r(y, 0) = \frac{1-2\delta}{1-\delta}$. Hence, $\max\{r : y \geq \pi_1 \geq \pi_2\} = r(y, 0) = \frac{1}{2}(1-\delta)y$ if $\delta \leq \frac{1}{2}$ and $\max\{r : y \geq \pi_1 \geq \pi_2\} = \frac{1}{2}(1-\delta)\left(2 - \sqrt{2(1-\delta)}\right)y\frac{\delta - 1 + \sqrt{2(1-\delta)}}{\delta\sqrt{2(1-\delta)}}$ if $\delta > \frac{1}{2}$ where the maximum is attained at $\pi_1 = \frac{1}{2\delta}\left(2 - \sqrt{2(1-\delta)}\right)y$ and $\pi_2 = 0$.

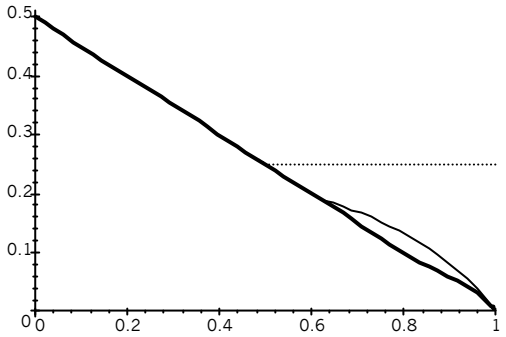
Given the above we can derive the conditions for when r is maximized at $\pi_1 = 1$ and $\pi_2 = 0$. If $\delta \leq \frac{1}{2}$ then we need $y \geq \delta$. If $\delta > \frac{1}{2}$ then we need $y \geq \delta$ and $\frac{1}{2}(1-\delta)\left(2 - \sqrt{2(1-\delta)}\right)y\frac{\delta - 1 + \sqrt{2(1-\delta)}}{\delta\sqrt{2(1-\delta)}} \leq \frac{1}{2}(1-\delta)$ so $\delta \leq y \leq \frac{\delta\sqrt{2(1-\delta)}}{\left(2 - \sqrt{2(1-\delta)}\right)\left(\delta - 1 + \sqrt{2(1-\delta)}\right)}$ which implies $y < 1$ and $\delta \leq 2\sqrt{2} - 2$.

When $\delta > 2\sqrt{2} - 2$ then

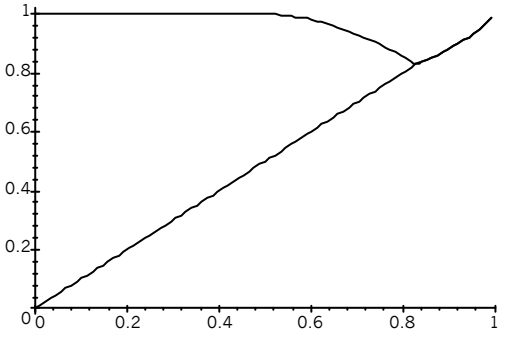
$$\begin{aligned} \sup r &= \max\left\{\frac{1}{2}(1-\delta), \frac{1}{2}(1-y), \frac{1}{2}(1-\delta)\left(2 - \sqrt{2(1-\delta)}\right)y\frac{\delta - 1 + \sqrt{2(1-\delta)}}{\delta\sqrt{2(1-\delta)}}\right\} \\ &= \frac{1}{2}(1-\delta)\frac{-4(1-\delta) + (3-\delta)\sqrt{2(1-\delta)}}{-4(1-\delta)^2 + \left(3 - 3\delta + \delta^2\right)\sqrt{2(1-\delta)}} \end{aligned}$$

which is attained if $y = \frac{\delta\sqrt{2(1-\delta)}}{-4(1-\delta)^2 + \left(3 - 3\delta + \delta^2\right)\sqrt{2(1-\delta)}} =: y^*$. So the censored reinforcement rule with cutoff level y^* attains minimax regret under deterministic payoffs among the

censored reinforcement rules. Below we graph the minimax regret value attainable among the censored reinforcement rules and include the value attained among the simple cutoff (thin line) and among the censored sampling rules. The relationships apparent from the graph are easily also verified formally.



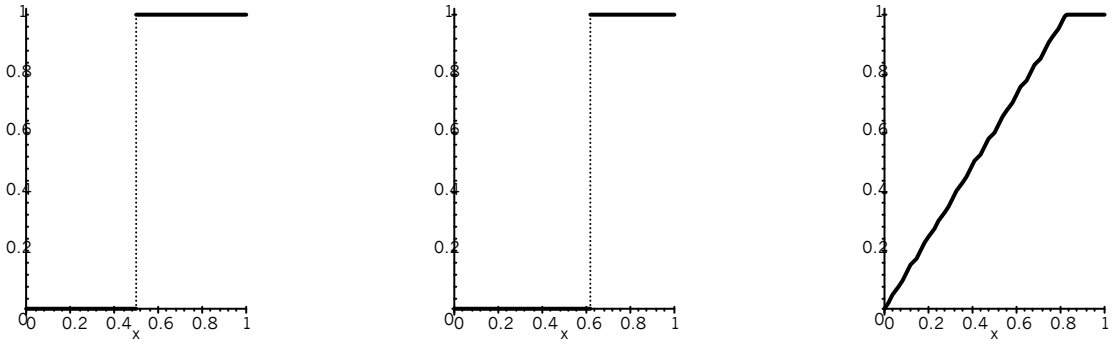
The figure below shows the upper and lower bound for the cutoff level y of the censored reinforcement rule in order to attain the maximal regret graphed as a thick line in the figure above. Upper and lower bound coincide when $\delta \geq 2\sqrt{2} - 2$ as the cutoff level y is uniquely determined in this range. Notice that in contrast to the previous rules analyzed there is no constraint on y from above when $\delta \leq \frac{1}{2}$.



When payoffs are random then we know less except for what we can derive (analogous to the analysis of censored sampling rules) from the fact that censored reinforcement rules and the grim trigger rules behave identically when payoffs are contained in $\{0, 1\}$.

Proposition 6 *Assume $n = 2$. (i) The censored reinforcement rule with cutoff level y attains minimax regret under deterministic payoffs if $\delta \leq y$ and either $\delta \leq \frac{1}{2}$ or $y \leq \frac{\delta\sqrt{2(1-\delta)}}{\left(2-\sqrt{2(1-\delta)}\right)\left(\delta-1+\sqrt{2(1-\delta)}\right)}$ and $\delta \leq 2\sqrt{2}-2$ (≈ 0.83). In particular, the censored reinforcement rule with cutoff level $2\sqrt{2}-2$ attains minimax regret under deterministic payoffs for all $\delta \leq 2\sqrt{2}-2$. (ii) If $\delta > 2\sqrt{2}-2$ then the censored reinforcement rule with cutoff level y^* attains minimax regret under deterministic payoffs among the censored reinforcement rules and achieves a strictly lower value of maximal regret than any simple cutoff or censored sampling rule. (iii) There is no censored reinforcement rule that attains minimax regret under random payoffs if $\frac{3}{2}-\frac{1}{2}\sqrt{5} < \delta \leq \frac{1}{2}\sqrt{5}-\frac{1}{2}$ (where $\frac{3}{2}-\frac{1}{2}\sqrt{5} \approx 0.38$).*

The next figures show the probabilities of choosing the first chosen action again as a function of the previous payoff x for the three rules above using the cutoffs that yield minimax regret under deterministic payoffs for the highest discount values: for the simple cutoff rule with cutoff level $y = \frac{1}{2}$, for the censored sampling rule with cutoff level $y = \frac{1}{2}\sqrt{5}-\frac{1}{2}$ (thick line) and for the censored reinforcement rule with cutoff level $y = 2\sqrt{2}-2$. These rules attain minimax regret under deterministic payoffs for discount values below $\frac{1}{2}$, $\frac{1}{2}\sqrt{5}-\frac{1}{2} \approx 0.62$ and $2\sqrt{2}-2 \approx 0.83$ respectively.



A.6 Sequential Minimax Regret

The decision maker obtains valuable information after his first choice when payoffs are deterministic. This information changes the set of possible decision problems he could be facing. In the following we search for behavior under which after each possible history the decision maker chooses a rule that minimizes maximal regret given his current information. For instance, after the decision maker has chosen each action at least once he will minimize

his maximal regret given this history by choosing forever an action that yielded the highest payoff. The simple cutoff rule with cutoff level $\frac{1}{2}$ does not have this property but nevertheless minimizes maximum regret from the ex-ante perspective before making the first choice if $\delta \leq \frac{1}{2}$. The reason being that for this range of discount values the maximal regret attainable against any decision problem is not changed if learning is improved once both actions have been chosen (which can take effect earliest in round 3).

In the following we will search for a rule that similar to subgame perfection attains minimax regret conditional on any history. Later we only consider deterministic payoffs and two actions only. In this case there are three different types of histories, namely before the first choice, after only haven chosen action i and finally after having chosen both actions. We say that a rule attains *sequential minimax regret (under deterministic payoffs)* if conditional on any such history it attains minimax regret under deterministic payoffs. More formally, h is called a *consistent history* if $h \in \cup_{m=0}^{\infty} \left(\{(i, \pi_i(D)), i = 1, .., n\}\right)^m$ for some deterministic decision problem D. So a history in which the same action yielded two different payoffs would not be consistent and is ruled out as such a history could not be generated by a deterministic decision problem. Let $\mathcal{D}(h)$ be the set of deterministic decision problems that can generate the consistent history h. Then f *attains minimax regret conditional on the consistent history* h if $f \in \arg\min_{f \in \mathcal{F}} \sup_{D \in \mathcal{D}(h)} \left(\max_{i=1,..,n} \pi_i(D) - \pi(D, f|h)\right)$ where $\pi(D, f|h)$ is the expected present value of discounted future payoffs yielded by rule f in decision problem D conditional on past history h. So formally f attains sequential minimax regret under deterministic payoffs if f attains minimax regret conditional on any consistent history h.

We proceed with our analysis analogously to the backwards induction algorithm. After both actions have been chosen at least once regret is minimized by choosing forever the action that attained the higher payoff. The value of regret is not influenced by what the rule prescribes if both actions are known to yield the same payoff.

Now assume that action i has been chosen in the first round. Let λ be the probability of choosing action i in round two under a rule that attains sequential minimax regret. Stationarity of the problem implies that we can assume that λ is the probability of choosing action i whenever only action i has been chosen in the past. Let v be the expected payoff then

$$v = \lambda\left((1-\delta)\pi_1 + \delta v\right) + (1-\lambda)\left((1-\delta)\pi_2 + \delta \max\{\pi_1, \pi_2\}\right)$$

so

$$v = \frac{1}{1-\delta\lambda}\lambda(1-\delta)\pi_1 + \frac{1}{1-\delta\lambda}(1-\lambda)((1-\delta)\pi_2 + \delta\max\{\pi_1,\pi_2\})$$

If $\pi_1 \geq \pi_2$ then

$$r = \pi_1 - v \leq \pi_1 - \frac{1}{1-\delta\lambda}\lambda(1-\delta)\pi_1 - \frac{1}{1-\delta\lambda}(1-\lambda)\delta\pi_1 = \frac{(1-\delta)(1-\lambda)\pi_1}{1-\delta\lambda}$$

and if $\pi_1 < \pi_2$ then

$$r = \pi_2 - \frac{1}{1-\delta\lambda}\lambda(1-\delta)\pi_1 - \frac{1}{1-\delta\lambda}(1-\lambda)\pi_2 = \frac{(1-\delta)\lambda(\pi_2-\pi_1)}{1-\delta\lambda} \leq \frac{(1-\delta)\lambda(1-\pi_1)}{1-\delta\lambda}.$$

Consequently minimax regret is attained if and only if λ solves $(1-\lambda)\pi_1 = \lambda(1-\pi_1)$ so $\lambda = \pi_1$.

Finally consider behavior in round 1. It is easily argued that the rule must prescribe to choose each action equally likely in the first round. Thus, any rule that attains sequential minimax regret behaves like the censored reinforcement rule with cutoff level 1 apart from possibly different behavior when it is known that both actions yield the same payoff. Recall from our analysis in Section A.5 that the censored reinforcement rule with cutoff level 1 attains minimax regret under deterministic payoffs if and only if $\delta \leq 1/2$. We summarize.

Proposition 7 *Assume $n = 2$. (i) If $\delta \leq 1/2$ then a rule attains sequential minimax regret if and only if it behaves like the censored reinforcement rule with cutoff level 1 except for when both actions are known to achieve the same payoff. (ii) There is no rule that attains sequential minimax regret for some $\delta > \frac{1}{2}$.*

Next we search for a rule that attains sequential minimax regret among the rules that are deterministic after round 1. So we must reconsider behavior after only action i has been chosen in the previous rounds where we assume w.l.o.g. that $i = 1$. The maximal regret when not switching equals $1-\pi_1$ while the regret to switching equals

$$\max\{\pi_1,\pi_2\} - ((1-\delta)\pi_2 + \delta\max\{\pi_1,\pi_2\}) = (1-\delta)(\max\{\pi_1,\pi_2\} - \pi_2) \leq (1-\delta)\pi_1$$

so minimax regret behavior prescribes to switch if $1-\pi_1 > (1-\delta)\pi_1$ which holds if and only if $\pi_1 < \frac{1}{2-\delta}$. Notice that behavior when $\pi_1 = \frac{1}{2-\delta}$ does not influence the value of maximal regret. Note that the censored sampling rule with cutoff level $\frac{1}{2-\delta}$ has the

desired behavior. Given our previous results in Proposition 5 this rule attains minimax regret under deterministic payoffs if and only if

$$\delta \leq \min\left\{\frac{1-\frac{1}{2-\delta}}{\frac{1}{2-\delta}}, \frac{1}{2-\delta}\right\} = \min\left\{1-\delta, \frac{1}{2-\delta}\right\}$$

which holds if and only if $\delta \leq 1/2$. Analogous to the case discussed above this yields a complete characterization of sequential minimax regret behavior among the rules that are deterministic after round 1.

Proposition 8 *Assume $n = 2$. (i) If $\delta \leq 1/2$ then a rule attains sequential minimax regret among the rules that are deterministic after round 1 if and only if it behaves like the censored sampling rule with cutoff level $1/(2-\delta)$ except after observing only action i that achieved a payoff $1/(2-\delta)$ and when both actions are known to achieve the same payoff. (ii) There is no rule that attains sequential minimax regret among the rules that are deterministic after round 1 for some $\delta > \frac{1}{2}$.*

It is interesting that (but not immediately apparent why) we obtain the same critical discount value in both propositions above.

Above we consider deterministic payoffs. In the following we briefly argue why there is no rule that attains sequential minimax regret under random payoffs when $n = 2$. Let $\mathcal{D}_{/}$ be the set of decision problems in which no action yields a deterministic payoff. Notice that the definition of minimax regret or sequential minimax regret does not change if we limit attention to decision problems in $\mathcal{D}_{/}$. Notice also that for any finite number of observations there is no decision problem $D \in \mathcal{D}_{\backslash}$ where the decision maker knows for sure that he is not facing D. Using our nomenclature above, all decision problems are consistent with any decision problem D in $\mathcal{D}_{\backslash}$. So the problem of finding a rule that attains minimax regret conditional on history h is equivalent to finding a minimax regret rule at the beginning of the game. We know from Schlag (2003) that any rule that attains minimax regret chooses each action equally likely in the first round. Consequently, a rule that attains sequential minimax regret under random payoffs chooses each action with probability $\frac{1}{2}$ in each round. Clearly this rule does not attain minimax regret under random payoffs and hence there is no rule that attains sequential minimax regret under random payoffs.

A.7 The Simple Reinforcement Rule

In the next two sections in which we no longer limit attention to two actions only we investigate two specific rules that involve only transition probabilities that are linear in payoffs. The rule considered in this section relies on the same minimal single round memory as the simple cutoff rule. It is called the *simple reinforcement rule* and prescribes the following behavior. Assign indices at random to the n actions. In the first round choose each action equally likely. From round two on choose the same action again with probability equal to x where x is the payoff received in the previous round. With probability $1-x$ switch to the action with the next higher index (modulo n).

Notice that the expected payoff of this rule in a general decision problem D is the same as in the deterministic decision problem D_0 in which $P_i(\pi_i(D), D_0) = 1$ for all i. Consequently regret only depends on the expected payoff of each action and hence minimax regret under deterministic payoffs yields the same conditions as minimax regret under random payoffs.

Assume that action i yields an expected payoff of π_i. Let w_i be the future discounted value from choosing action i in round one. Then $w_i = (1-\delta)\pi_i + \pi_i \delta w_i + (1-\pi_i)\delta w_{(i+1) \bmod n}$ and it is straightforward to verify that

$$w_i = (1-\delta)\frac{\sum_{k=i}^{n+i-1}\delta^{k-i}\left(\prod_{j=i}^{(k-1) \bmod n}(1-\pi_j)\right)\pi_{k \bmod n}\prod_{j=(k+1) \bmod n}^{(i-1) \bmod n}(1-\delta\pi_j)}{\prod_{j=1}^{n}(1-\delta\pi_j) - \delta^n\prod_{j=1}^{n}(1-\pi_j)}.$$

Clearly $\frac{1}{n}\sum_{i=1}^{n} w_i$ is then the future discounted value of using the simple reinforcement rule.

If $\pi_1 = 1$ and $\pi_i = 0$ for $i \geq 2$ then $w_1 = 1$ and $w_i = \delta^{n+1-i}$ for $i \geq 2$ so $\frac{1}{n}\sum_{i=1}^{n} w_i = \frac{1}{n(1-\delta)}(1-\delta^n)$ so $r = 1 - \frac{1}{n(1-\delta)}(1-\delta^n)$. In particular, $n = 2$ yields $r = \frac{1}{2}(1-\delta)$ and $n = 3$ yields $\frac{1}{3}\left(2-\delta-\delta^2\right)$. While we allow for general n in our description of behavior the complexity of the analysis only allows us to state a result when $n \leq 3$.

Proposition 9 *Consider either deterministic or random payoffs. (i) If $n = 2$ then the simple reinforcement rule attains minimax regret if $\delta \leq \sqrt{2}-1$ (≈ 0.41).*[5] *Maximal regret equals $\frac{1}{2}(1-\delta)$. (ii) If $n = 3$ then the simple reinforcement rule attains minimax regret if $\delta \leq \frac{1}{2}\sqrt{4\sqrt{3}-3} - \frac{1}{2}$ (≈ 0.49). Maximal regret equals $\frac{1}{3}\left(2-\delta-\delta^2\right)$.*

[5] Moreover, Schlag (2003) shows that there the is no rule with single round memory that attains minimax regret under random payoffs for $\delta > \sqrt{2}-1$.

Proof. The proof for $n = 2$ is contained in (Schlag, 2003). Consider $n = 3$. It is easily verified for the simple reinforcement rule that

$$r = \pi_1 - \frac{\begin{array}{c}\frac{1}{3}(\pi_1+\pi_2+\pi_3) + \left(\frac{1}{3}\pi_1+\frac{1}{3}\pi_2+\frac{1}{3}\pi_3-\pi_1\pi_2-\pi_1\pi_3-\pi_2\pi_3\right)\delta \\ +\frac{1}{9}(1-(1-3\pi_1)(1-3\pi_2)(1-3\pi_3))\delta^2\end{array}}{\left(1+(1-\pi_1-\pi_2-\pi_3)\delta+(1-\pi_1-\pi_2-\pi_3+\pi_1\pi_2+\pi_1\pi_3+\pi_2\pi_3)\delta^2\right)},$$

that $r = \frac{1}{3}(1-\delta)(2+\delta)$ if $\pi_1 = 1$ and $\pi_2 = \pi_3 = 0$, that $\frac{d}{d\pi_2}r \leq 0$ and that $\pi_2 = 0$ implies $\frac{d}{d\pi_3}r \leq 0$. If $\delta \leq \frac{1}{2}\sqrt{4\sqrt{3}-3} - \frac{1}{2} \approx 0.49098$ then we also obtain that $\pi_2 = \pi_3 = 0$ implies $\frac{d}{d\pi_1}r \geq 0$. Consequently, $\delta \leq \frac{1}{2}\sqrt{4\sqrt{3}-3} - \frac{1}{2}$ implies that r is maximized if $\pi_1 = 1$ and $\pi_2 = \pi_3 = 0$.

Similarly it follows immediately that the single reinforcement rule is a best response under the prior that puts equal weight on $\{D_i, i = 1, .., 3\}$ where D_i is the decision problem in which $P_i(1) = 1$ and $P_j(1) = 0$ for $j \neq i$.

The fact that the simple reinforcement rule is a best response against this prior and that this prior maximizes regret of the rule when $\delta \leq \frac{1}{2}\sqrt{4\sqrt{3}-3} - \frac{1}{2}$ implies that f attains minimax regret (Schlag, 2003). ■

When we use this rule in the main section we will need to know how this rule performs in the long run. Let q_i be the average number of rounds in which action i is chosen in the long run so $q_i = \pi_i q_i + \left(1 - \pi_{(i-1) \bmod n}\right) q_{i-1}$ and hence

$$\begin{aligned} q_i &= \frac{1}{(1-\pi_i)\sum_{k=1}^{n}\frac{1}{1-\pi_k}} \text{ if } \max_k\{\pi_k\} < 1 \\ q_i &= \frac{1}{\#\{k : \pi_k = 1\}} \text{ if } \pi_i = 1 \\ q_i &= 0 \text{ if } \pi_i < 1 = \max_k\{\pi_k\} \ . \end{aligned}$$

A.8 The Two State Confidence Rule

Next consider the following rule we call the *two state confidence rule*. This rule requires two rounds of memory. For each action there are two states which can be interpreted as confidence levels low and high. As above, assign indices at random to the n actions and choose each action equally likely in the first round. Enter the low confidence state of the respective action. After being in the low confidence state of a given action, with probability equal to the payoff x received in the previous round choose the same action again and transit to the high state. Otherwise, i.e. with probability $1 - x$, choose the

action with the next higher index (modulo n) and enter its low confidence state. Thus you never are in the low confidence state of the same action is in two consecutive rounds. After being in the high confidence state of a given action, choose the same action again in the next round. Remain in the high state with probability x and transit to the low confidence state of the same action with probability $1-x$.

Let l_i and h_i be the values of being in the low and high states respectively of action i for $i = 1, .., n$. Then

$$\begin{aligned} l_i &= (1-\delta)\,\pi_i + \pi_i \delta h_i + (1-\pi_i)\,\delta l_{(i+1) \bmod n} \\ h_i &= (1-\delta)\,\pi_i + \pi_i \delta h_i + (1-\pi_i)\,\delta l_i \end{aligned}$$

so

$$h_i = \frac{(1-\delta)\,\pi_i + \delta\,(1-\pi_i)\,l_i}{1-\pi_i\delta}$$

and hence $l_i = a_i + b_i l_{(i+1) \bmod n}$ where

$$a_i = \frac{(1-\delta)\,\pi_i}{1-\pi_i\delta-\pi_i\delta^2+\pi_i^2\delta^2} \text{ and } b_i = \frac{\delta\,(1-\pi_i-\delta\pi_i+\delta\pi_i^2)}{1-\pi_i\delta-\pi_i\delta^2+\pi_i^2\delta^2}\ .$$

Given the above it is easy to verify that

$$l_i = \frac{\sum_{k=i}^{(n+i-1) \bmod n} a_k \prod_{j=i}^{k-1} b_j}{1-\prod_{k=1}^{n} b_k}\ .$$

Notice that $r = \max \pi_i - \frac{1}{n}\sum_{i=1}^{n} l_i$, e.g. if $n = 2$ then

$$r = \frac{1}{2}\frac{\left(1+\delta-2\delta\pi_1-2\delta^2\pi_1+2\delta^2\pi_1^2\right)(\pi_1-\pi_2)}{\delta^2\,(1-\pi_1)^2+\delta^2\,(1-\pi_2)^2+(1-\delta)\,(1+\delta\,(2-\pi_1-\pi_2))} \text{ if } \pi_1 \geq \pi_2.$$

For long run behavior let q_i^l and q_i^h be the average number of rounds in which the low and high state respectively of action i are visited. Then $q_i^l + q_i^h$ is the average number of rounds in which action i is chosen. So $q_i^l = (1-\pi_i)\,q_i^h + \left(1-\pi_{(i-1) \bmod n}\right) q_{(i-1) \bmod n}^l$ and $q_i^h = \pi_i\left(q_i^l + q_i^h\right)$ which is has the unique solution

$$q_i^l = \frac{1}{(1-\pi_i)\sum_{j=1}^{n}\frac{1}{(1-\pi_j)^2}} \text{ and } q_i^h = \frac{\pi_i}{(1-\pi_i)^2\sum_{j=1}^{n}\frac{1}{(1-\pi_j)^2}}$$

and hence

$$q_i = \frac{1}{(1-\pi_i)^2\sum_{j=1}^{n}\frac{1}{(1-\pi_j)^2}}\ .$$

Again we are only able to present results for $n \leq 3$.

Proposition 10 *Consider either deterministic or random payoffs and* $n \in \{2,3\}$. *The two state confidence rule attains minimax regret for any* $\delta \leq \frac{1}{2}\sqrt{5} - \frac{1}{2}$ (≈ 0.62).

Proof. The proof for $n = 2$ is contained in (Schlag, 2003). Consider $n = 3$. It is clear that the two state confidence rule is a best response against the prior shown in the proof of the above proposition. All that needs to be shown is that $\pi_1 = 1$ and $\pi_2 = \pi_3 = 0$ maximizes regret of this rule when $\delta \leq \frac{1}{2}\sqrt{5} - \frac{1}{2}$. Here the proof is a bit more messy than in the case of the simple reinforcement rule. The analysis is straightforward when the steps below are followed.

Above we already showed how to calculate r. The first part of the proof is to show that $\frac{d}{d\pi_3} r \leq 0$ and hence that r is maximized by setting $\pi_3 = 0$.

$\frac{d}{d\pi_3} r$ can be written as $-\gamma\beta\alpha^2$ where

$$\begin{aligned}\gamma &= 1 + (1 - 2\pi_2 - \pi_1)\,\delta + \left(\pi_1^2 + 2\pi_2^2 - 3\pi_1 + 3\pi_1\pi_2 + 1 - 3\pi_2\right)\delta^2 \\ &\quad + (1 - \pi_1 - \pi_2)\,(-2\pi_1 + 3\pi_1\pi_2 - \pi_2)\,\delta^3 + 3\pi_1\pi_2\,(1 - \pi_1)\,(1 - \pi_2)\,\delta^4\end{aligned}$$

and

$$\begin{aligned}\beta &= 1 + (1 - 2\pi_1 - \pi_2)\,\delta + \left(1 - 3\pi_1 + \pi_1^2 - 3\pi_2 + \pi_2^2 - \pi_3^2 + 3\pi_1\pi_2 + 2\pi_1\pi_3\right)\delta^2 \\ &\quad + \begin{pmatrix} -\pi_1 + \pi_1^2 - 2\pi_2 + \pi_2^2 - \pi_3^2 + 6\pi_1\pi_2 - 2\pi_1\pi_2^2 \\ -2\pi_1^2\pi_2 + \pi_1\pi_3^2 + 2\pi_2\pi_3 + \pi_2\pi_3^2 - 4\pi_1\pi_2\pi_3 \end{pmatrix}\delta^3 \\ &\quad + \begin{pmatrix} 3\pi_1\pi_2 - 2\pi_1^2\pi_2 - 2\pi_1\pi_2^2 + \pi_1^2\pi_2^2 + \pi_1\pi_3^2 - \pi_1^2\pi_3^2 \\ +\pi_2\pi_3^2 - \pi_2^2\pi_3^2 + 2\pi_1^2\pi_2\pi_3 + 2\pi_1\pi_2^2\pi_3 - 4\pi_1\pi_2\pi_3 \end{pmatrix}\delta^4 .\end{aligned}$$

$\gamma \geq 0$ follows from the fact that $\frac{d}{d\pi_1}\gamma \leq 0$ and $\gamma \geq 0$ if $\pi_1 = 1$.

Since β is concave in π_3 it is sufficient to verify $\beta \geq 0$ for $\pi_3 = 0$ and $\pi_3 = 1$. For $\pi_3 \in \{0,1\}$ it is easily verified that $\frac{d}{d\pi_2}\frac{d}{d\pi_2}\beta \geq 0$ and that $\frac{d}{d\pi_2}\beta \leq 0$ holds for $\pi_2 = 1$ which means that $\frac{d}{d\pi_2}\beta \leq 0$ for all π_2. Moreover it can be shown that $\beta \geq 0$ holds if $\pi_2 = 1$ which means that $\beta \geq 0$ for all π_2. Here we need that $\delta \leq \frac{1}{2}\sqrt{5} - \frac{1}{2}$ holds in order to prove the statements regarding $\pi_3 = 0$.

Thus $\frac{d}{d\pi_3} r \leq 0$ so r is maximized by setting $\pi_3 = 0$.

Next we show that r is maximized by setting $\pi_2 = 0$.

Given $\pi_3 = 0$ we find that $\frac{d}{d\pi_2} r$ can be written as $-\sigma\omega$ where $\omega \geq 0$ and

$$\sigma = 1 + (1 - \pi_1)\,\delta + \left(1 - 3\pi_1 + \pi_1^2 - \pi_2^2\right)\delta^2 + \left(\pi_1^2 - 2\pi_1 - \pi_2^2 + 2\pi_1\pi_2 + \pi_1\pi_2^2\right)\delta^3 + \pi_1\pi_2^2\,(1 - \pi_1)\,\delta^4$$

where σ is concave in π_2 and $\pi_2 \in \{0,1\}$ implies $\frac{d}{d\pi_1}\sigma \leq 0$ and $\sigma \geq 0$ if $\pi_1 = 1$. Thus $\frac{d}{d\pi_2}r \leq 0$ and r is maximized when $\pi_2 = 0$.

Finally we show that $\frac{d}{d\pi_1}r \geq 0$. Given $\pi_2 = \pi_3 = 0$ we find that $\frac{d}{d\pi_1}r = \mu\eta^2/3$ where

$$\begin{aligned}\mu &= 2 + 2(2 - 3\pi_1)\delta + \left(6 - 18\pi_1 + 10\pi_1^2\right)\delta^2 + 2(1 - \pi_1)\left(2 - 10\pi_1 + 3\pi_1^2\right)\delta^3 \\ &\quad + \left(2 - 18\pi_1 + 32\pi_1^2 - 18\pi_1^3 + 3\pi_1^4\right)\delta^4 + \pi_1\left(-6 + 19\pi_1 - 18\pi_1^2 + 6\pi_1^3\right)\delta^5 + 3\pi_1^2(1 - \pi_1)^2\delta^6.\end{aligned}$$

It can be verified that μ is concave in π_1 as $\left(\frac{d}{d\pi_1}\right)^3\mu \leq 0$ and $\left(\frac{d}{d\pi_1}\right)^2\mu \geq 0$ if $\pi_1 = 1$. Moreover $\mu \geq 0$ holds for $\pi_1 = 0$ and also holds for $\pi_1 = 1$ provided $\delta \leq 0.7$. Thus $\frac{d}{d\pi_1}r \geq 0$ and r is maximized when $\pi_1 = 1$ and $\pi_2 = \pi_3 = 0$.

Since the two state confidence rule is a best response to uniform prior over decision problems in which one action yields 1 and the other two yield 0 (see proof of Proposition 9) we obtain that it attains minimax regret. ■