

On the Robustness of Strategic Experimentation to Persuasive Cheap Talk*

Jonathan Rosborough

Department of Economics, St. Francis Xavier University, Antigonish, Canada
Email: rosboro@hotmail.com

Received July 24, 2012; revised August 26, 2012; accepted September 27, 2012

ABSTRACT

This paper develops a model in which a privately informed seller attempts to indirectly influence the experimentation strategy of a buyer by sending costless signals. The question under consideration is whether there is any credible way in which this single rational seller could influence the buyer's decisions. We provide bounds on information transmission in equilibrium, and show that there exists no reporting strategy for the seller which changes the experimentation strategy of the buyer. These results demonstrate the robustness of a class of learning models to coercion.

Keywords: Learning; Experimentation; Persuasion; Cheap Talk; Multi-Armed Bandits

1. Introduction

Much of the existing literature on learning and uncertainty centers on the case of an individual decision maker choosing sequentially among a fixed set of alternatives. In many economic situations, these alternatives are provided by agents whose welfare is affected by the choices made by the decision maker. Economic reasoning suggests that such an agent would have a vested interest in influencing the decision maker's experimentation behaviour.

In this paper, we develop a model in which a privately informed seller attempts to indirectly influence the experimentation strategy of a buyer by sending costless signals. The question under consideration is whether there is any credible way in which this single rational seller could influence the buyer's decisions. We demonstrate that although there may be some information transmission in equilibrium, there exists no reporting strategy for the seller which changes the experimentation strategy of the buyer. A careful examination of this negative result points to numerous potentially fruitful directions for future work that are discussed in closing.

The organization of the paper is as follows. Subsection 1.1 discusses some of the relevant literature on both sequential learning and strategic information transmission. Section 2 outlines the basic framework of the model and presents the main results. Section 3 concludes the paper with a discussion of avenues for further research.

Related Literature

The proposed model builds on several separate strands in the economics of information literature. The first area deals with sequential learning and experimentation models known as multi-armed bandit problems. A bandit problem involves sequential selections from a number of stochastic processes (or "arms") which have unknown characteristics so learning can take place as the processes are observed. Until Rothschild's [1] contribution, it was considered innocuous to assume that a decision maker in a situation involving uncertainty knew all relevant parameters of the stochastic distributions of interest. Rothschild's central result was that if these distributions are not known, then there is nothing to guarantee even in the long run that correct decisions (*i.e.* choosing the "best" arm) will occur through experimentation. Given the possibility of such persistent "mistakes" made by experimenters, a natural direction in the literature has been to give the arms a strategic/competitive role and examine if this inefficiency is exploited or eliminated through competition. Work in this area, particularly Bergemann and Välimäki [2], and Bar-Isaac [3], is of direct relevance to the model presented here.

Bergemann and Välimäki model a situation in which a single consumer buys a stream of goods of initially unknown quality from different sellers over time. The consumer learns about product quality through experimentation while sellers affect the cost of experimentation through price competition. However, unlike this model in which experimentation is affected directly through manipulation of the cost of successive trials, the model pre-

*I am grateful to Braz Camargo, Igor Livshits, Chris Bennett and Christopher Hajzler and participants at the UWO Theory Workshop for useful comments and suggestions. All errors are mine.

sented here exogenizes these costs and allows only an indirect role of the sellers through signalling. Our model is closely related to Bar-Isaac's paper although in his model, sellers signal buyers through production decisions while we consider these signals as costless reports sent by the sellers.

The notion that buyers' decisions can be affected by costless claims made by privately informed sellers is not new. Following Crawford and Sobel [4], cheap talk models have been applied to a variety of buyer-seller environments. While the model in Crawford and Sobel is a one shot model, the basic framework remains the same in a repeated setting¹. In the canonical cheap talk model, a sender observes a signal (his "type") and then sends a message to a receiver who takes an action that determines the payoffs of both agents. As will be explained below, the model in this paper differs from the basic cheap talk model in two fundamental ways. First, the signal observed by the seller is not a perfect indication of his "type". That is, the seller only receives noisy information about the true quality of the product he is selling to the buyer. Second, after the buyer makes her decision, her reward provides only partial information about the seller's type. Both of these assumptions together allow for a dynamic interaction of the sender and receiver with the potential for incomplete learning on the part of the buyer.

Also of relevance to the work in this paper is the model of *reputational cheap talk* in Ottaviani and Sørensen [6]. In reputational cheap talk, an expert (sender) gets a private noisy signal about the state of the world and sends a forecast to an evaluator (receiver). The informativeness of the signal received by the expert depends on his/her ability and the evaluator uses the forecast and the realized state of the world to form a belief about this ability of the expert. The model presented here shares a similar information structure to Ottaviani and Sørensen but instead focuses on the sender's role as a seller of a commodity with an uncertain payoff distribution, instead of an expert with preferences for esteem conferred by reputation.

Perhaps most closely related to the model presented here is that of insider information in Benabou and Laroque [7]. In their model, a market insider (or guru) receives a private signal about the likelihood that a particular asset will pay off a positive reward in that period. The insider, who may be truthful or strategic, then sends a signal to the market about the expected payoff of the asset and has the ability to engage in post announcement speculation. The concern of a strategic insider is thus a tradeoff between long run gains from building influential credibility and short run gains from market manipulation.

¹See Aumann and Hart [5].

A critical assumption in this insider model is that the distribution of returns for the asset is common knowledge. Thus, the only learning that takes place is about whether the insider is truthfully reporting his private advance information about whether the asset will succeed or fail in a given period. In the model presented below, we do not assume that the reward distribution of the asset is known. As such, we model learning as updating beliefs about the quality of the asset and examine whether a seller can credibly commit to a reporting strategy that can affect those beliefs.

2. Model

We consider a finite horizon discrete time model in which time is indexed by $t = 0, \dots, T$. There are two agents, a buyer (B) and a seller (S) and two assets X and Y . The assets pay off 1 or 0 (success or failure) with the following probabilities:

$$P_x(\text{success}) = 1 - P_x(\text{failure}) = x$$

$$P_y(\text{success}) = 1 - P_y(\text{failure}) = y$$

The rewards of 1 or 0 go to the buyer in the event that the chosen asset succeeds or fails and the seller receives a fixed reward that depends only on the asset chosen by the buyer.

2.1. Information Structure

Information and Timing in the model is as follows. The success probability x is commonly known to be x_0 . The success probability y is known to take only one of two values $y = y_0 < x_0$ or $y = y_1 > x_0$. The common, non-degenerate prior that $y = y_1$ is $\rho_0^S = \rho_0^B = \rho_0$.

At the beginning of each period, S receives a noisy private signal $w \in W$ about the success probability y where W is a compact subset of \mathbb{R} . The signal S receives is drawn from the continuous conditional probability density function $f(w|y)$ with full support. We assume that $f(\cdot|y)$ is common knowledge. After observing his private signal w , S then sends a message $m \in M$ to B via a reporting strategy $\varphi: W \rightarrow P(m)$ which specifies the conditional probability of sending message m upon receiving signal w . After observing the message, B then processes the information in m and chooses her action $a_t \in \{0, 1\}$ for that period which consists of choosing either asset X ($a_t = 0$) or Y ($a_t = 1$).

The success or failure of the chosen asset is publicly observed and beliefs about the success probability y are updated. We assume that the trial outcome and the signal observed by the seller are independent conditional on y . The timing and structure in each period is identical.

2.2. Payoffs and Preferences

Per period rewards depend on the action a chosen by B and the realization of the success or failure of the chosen asset. The seller receives an amount $c > 0$ if the buyer chooses Y in period t . Thus, the rewards to the seller are:

$$r^S(a_t) = ca_t$$

The rewards to the buyer are:

$$r^B(a_t, n_t) = \begin{cases} 1, & \text{if chosen asset succeeds} \\ 0, & \text{if chosen asset fails} \end{cases}$$

We assume that the buyer and seller have discount factors β^i and choose their strategies to maximize the expected discounted reward streams:

$$U^S(a) = E \left[\sum_{t=0}^T \beta^t r^S(a_t) \middle| w, \rho^S \right]$$

$$U^B(a, n) = E \left[\sum_{t=0}^T \beta^t r^B(a_t, n_t) \middle| m, \rho^B \right]$$

2.3. Belief Updating

Recall that beliefs for B and S in each period t are real numbers $\rho_t^i \in [0, 1], i \in \{S, B\}$, representing the probabilities they place on the state $y = y_1$. The seller is effectively at an informational advantage in the sense that although both B and S observe the outcome of the chosen asset, the seller also uses his private information to update beliefs, while the buyer can use only the part of this private information that can be credibly communicated to her.

Upon receiving the signal $w \in W$, the seller updates his beliefs according to the rule:

$$\hat{\rho}_t^S = \frac{\rho_t^S f(w|y_1)}{\rho_t^S f(w|y_1) + (1 - \rho_t^S) f(w|y_0)} \tag{1}$$

We refer to such updating by the seller as “signal updating”. Given these updated beliefs, he send the message m to B . Now we must specify two forms of belief updating for B , how beliefs are updated following a message from S and updating following a success/failure observation of an asset. We refer to the former as “message updating” and the latter as “trial updating”. Since there is in effect double updating, in order to align time subscripts, we denote an updated

belief in time t :

- 1) $\hat{\rho}_t^B$ following a message from S . (*Message updating*);
- 2) ρ_{t+1}^B following a success/failure observation of an asset. (*Trial updating*).

First, in order to use the message m to update beliefs, B must form a conjecture $\hat{\phi}^t$ about the reporting strategy of S in period t . Given this conjecture, a message m_t in period t results in updated beliefs: (see Equation (2)).

Next, we examine how beliefs are updated following a trial of each asset. Recall that both B and S engage in such trial updating. If the known arm X is chosen then no information regarding the unknown arm Y is obtained. Thus in this case clearly $\rho_{t+1}^B = \rho_{t+1}^S = \hat{\rho}_t$. If the unknown arm is chosen, then updated beliefs will depend on the outcome of the trial:

$$\rho_{t+1}^i = \begin{cases} \frac{\hat{\rho}_t^i y_1}{\hat{\rho}_t^i y_1 + (1 - \hat{\rho}_t^i) y_0} & \text{if trial is a success} \\ \frac{\hat{\rho}_t^i (1 - y_1)}{\hat{\rho}_t^i (1 - y_1) + (1 - \hat{\rho}_t^i) (1 - y_0)} & \text{if trial is a failure} \end{cases} \tag{3}$$

for $i \in \{S, B\}$.

To summarize, since the seller observes both the signal w and the success/failure of the chosen asset, his beliefs are updated according to (1) and (3). Since the buyer observes the message m and the trial outcome, her beliefs are updated according to (2) and (3). We are now ready to define an equilibrium for this game.

Definition 1. *A Perfect Bayesian Nash Equilibrium for this game is a reporting strategy for S which is a sequence of contingent reporting rules*

$\{\varphi^t\}_{t=0}^T$; *a sequence of contingent decisions for B*

$\{a_t\}_{t=0}^T$; *and a sequence of conjectures $\{\hat{\phi}^t\}_{t=0}^T$; such*

that in each period t :

- 1) if m^* is in the support of $\varphi^t(\cdot|w)$, then

$$m^* \in \arg \max_{m \in M} E_t \left[\sum_{s=t}^T \beta^{s-t} r^S(a_s(m)) \middle| w, \rho_t^S, \hat{\rho}_t^B \right]$$

- 2) for each message m ,

$$a_t \in \arg \max_{a \in \{0,1\}} E_t \left[\sum_{s=t}^T \beta^{s-t} r^B(a_s, n_s) \middle| a_t, m, \hat{\rho}_t^B \right].$$

- 3) the conjecture $\hat{\phi}^t$ is consistent: $\hat{\phi}^t = \varphi^t$

$$\hat{\rho}_t^B = \frac{\rho_t^B \int_w \hat{\phi}^t(m_t|w) f(w|y_1) dw}{\rho_t^B \left[\int_w \hat{\phi}^t(m_t|w) f(w|y_1) dw \right] + (1 - \rho_t^B) \left[\int_w \hat{\phi}^t(m_t|w) f(w|y_0) dw \right]} \tag{2}$$

4) Beliefs are updated via Bayes' Rule. B 's beliefs updated according to (2) and (3). S 's beliefs updated according to (1) and (3).

With this definition in place we now turn our attention to equilibrium behaviour of the buyer and seller.

2.4. Characterizing Equilibria

In order to characterize the equilibria of this model we begin by solving the model backwards beginning in period T . For the analysis that immediately follows, let us suppose that all actions, rewards, and updating, from period $T-1$ have occurred. Thus, the buyer enters period T with a belief ρ_T^B that $y = y_1$. Observe that since there is no opportunity for learning in the last period, the buyer behaves myopically in the sense that after observing the message m , she maximizes her period T expected payoff given $\hat{\rho}_T$. This implies that after message updating has occurred in period T , the buyer will choose Y if and only if

$$\hat{\rho}_T^B y_1 + (1 - \hat{\rho}_T^B) y_0 \geq x_0$$

or

$$\hat{\rho}_T^B \geq \frac{x_0 - y_0}{y_1 - y_0} \tag{4}$$

We will refer to the value of $\hat{\rho}_T^B$ for which (4) holds with equality as $\bar{\rho}_T$. We now turn to the strategy of the seller. Clearly, the optimal strategy of the seller will depend not only on the signal w he observes but also on the beliefs of the buyer. The question is whether there is any credible reporting rule such that the seller could affect the buyer's period T decision. The following proposition demonstrates that there is not.

Proposition 1. *There exists no equilibrium reporting rule φ^T such that the message m affects the decision a_T of the buyer.*

Proof. We first observe that if ρ_T^B is such that for all w :

$$\frac{\rho_T^B f(w|y_1)}{\rho_T^B f(w|y_1) + (1 - \rho_T^B) f(w|y_0)} \geq \bar{\rho}_T$$

or

$$\frac{\rho_T^B f(w|y_1)}{\rho_T^B f(w|y_1) + (1 - \rho_T^B) f(w|y_0)} < \bar{\rho}_T$$

then there is no signal that could affect the decision of

the buyer even if it were truthfully reported. In other words, ρ_T^B is either sufficiently low such that there is no news good enough to convince her to choose Y , or sufficiently high so that there is no news bad enough to convince her to choose X . In this case any reporting strategy for the seller is credible (even truth-telling) but we still have that a_T is a trivial function of m .

Clearly in equilibrium S would prefer to transmit no information (*i.e.* send one message for every signal w) rather than transmit information that would induce the buyer to choose X . As such, let us consider the case where $\rho_T^B < \bar{\rho}_T$ and where there exists some signal w_l such that for all signals $w \geq w_l$

$$\frac{\rho_T^B f(w|y_1)}{\rho_T^B f(w|y_1) + (1 - \rho_T^B) f(w|y_0)} \geq \bar{\rho}_T$$

Since W is a compact subset of \mathbb{R} , there is some signal \bar{w} such that $\bar{w} = \max\{w|w \in W\}$. Notice that all signals $w \in [w_l, \bar{w}]$ (if reported truthfully) would result in positive updating and so it must be the case that for all such signals

$$\frac{f(w|y_1)}{f(w|y_0)} > 1 \Rightarrow f(w|y_1) - f(w|y_0) > 0 \tag{5}$$

we now show that there is no credible reporting strategy in which the seller can truthfully communicate any subset of these signals. To begin, suppose that there exists a reporting rule φ such that there is a set of signals $\omega \subseteq [w_l, \bar{w}]$ and a set of messages $\mu \subset M$ with

$$\int_{\omega} \varphi(m|w) = \begin{cases} 1 & \text{For all } m \in \mu \\ 0 & \text{Otherwise} \end{cases} \tag{6}$$

and

$$\int_{\sim\omega} \varphi(m|w) = \begin{cases} 1 & \text{For all } m \in \sim\mu \\ 0 & \text{Otherwise} \end{cases} \tag{7}$$

where

$$\sim\mu \equiv \{m \in M | m \notin \mu\} \text{ and } \sim\omega \equiv \{w \in W | w \notin [w_l, \bar{w}]\}.$$

Given this reporting rule, recall that B 's updated beliefs following any message $m \in \sim\mu$ would be (see Equation (8)).

Thus, at a minimum, for the seller's reporting rule to be credible we must have for all $m \in \sim\mu$:

$$\int_w \varphi(m|w) [f(w|y_1) - f(w|y_0)] dw > 0$$

$$\hat{\rho}_T^B = \frac{\rho_T^B \int_w \varphi(m|w) f(w|y_1) dw}{\rho_T^B \int_w \varphi(m|w) f(w|y_1) dw + (1 - \rho_T^B) \left[\int_w \varphi(m|w) f(w|y_0) dw \right]} \tag{8}$$

so

$$\int \int_{-\mu}^W \varphi(m|w) [f(w|y_1) - f(w|y_0)] dw dm > 0$$

adding $\int \int_{\mu}^W \varphi(m|w) [f(w|y_1) - f(w|y_0)] dw dm$ to both sides gives us

$$\int \int_{\mu}^W \varphi(m|w) [f(w|y_1) - f(w|y_0)] dw dm < 0$$

$$\text{Since } \int \int_{\mu}^W \left| \varphi(m|w) [f(w|y_1) - f(w|y_0)] \right| dw dm < \infty,$$

we can apply Fubini's theorem to rewrite this expression as

$$\int_{\mu}^W \varphi(m|w) [f(w|y_1) - f(w|y_0)] dw dm < 0$$

or

$$\int_{\omega} [f(w|y_1) - f(w|y_0)] dw < 0$$

but (5) implies $\int_{\omega} [f(w|y_1) - f(w|y_0)] dw > 0$.

Given Proposition 1, which characterize optimal strategies in period T , we may now look at the equilibria of the entire game. To begin, consider the problem of the buyer in period $T-1$ deciding between X or Y after she has processed the information in S 's period $T-1$ message. Observe that since the seller cannot affect the probability that $\hat{\rho}_T^B \geq \bar{\rho}_T$ we know that this probability is an increasing function of $\hat{\rho}_{T-1}^B$. In other words, for a given $\hat{\rho}_{T-1}^B$, the distribution of possible updated beliefs in period T first order stochastically dominates the distribution of possible updated beliefs for any $\rho < \hat{\rho}_{T-1}^B$. It is well known that this condition implies that the optimal strategy for the seller in period $T-1$ is a cutoff strategy². That is, if $a_{T-1} = 1$ is optimal for a belief $\hat{\rho}_{T-1}^B$, then $a_{T-1} = 1$ is optimal for any belief greater than $\hat{\rho}_{T-1}^B$. Thus, there exists some $\bar{\rho}_{T-1}$ such that the optimal decision of the buyer in period $T-1$ is

$$a_{T-1} = \begin{cases} 1 & \text{if } \hat{\rho}_{T-1}^B \geq \bar{\rho}_{T-1} \\ 0 & \text{Otherwise} \end{cases} \quad (9)$$

Observe that given the cutoff belief strategy of the buyer in period $T-1$, the seller's discounted expected payoff in period $T-1$ is an increasing function of $\hat{\rho}_{T-1}^B$. As such, the seller would always weakly prefer sending no information to sending a message that would reduce $\hat{\rho}_{T-1}^B$. Also, if the seller's discounted expected payoff is strictly increasing in $\hat{\rho}_{T-1}^B$ (i.e. at $\bar{\rho}_{T-1}$ and the point where $a_T = 1$ regardless of the trial outcome in $T-1$) the seller would strictly prefer to sending no information

to sending a message that would reduce $\hat{\rho}_{T-1}^B$. Thus we have the following proposition.

Proposition 2. *In period $T-1$, the seller cannot credibly transmit any information which would strictly increase his expected discounted payoff.*

Proof. In period $T-1$, there are at most two regions of $\hat{\rho}_{T-1}^B$ in which information transmission could lead to a strict increase in the expected payoff of the seller. The first is the region in which $\hat{\rho}_{T-1}^B < \bar{\rho}_{T-1}$ and where there exists a set of signals $[w_l, \bar{w}]$ that would induce $\hat{\rho}_{T-1}^B \geq \bar{\rho}_{T-1}$ if they could be honestly reported. In this case, the argument in proposition 1 shows that there is no credible reporting strategy φ^{T-1} that could induce $\hat{\rho}_{T-1}^B \geq \bar{\rho}_{T-1}$.

The only other region in which information transmission could lead to a strict increase in the expected payoff of the seller is where $\hat{\rho}_{T-1}^B \geq \bar{\rho}_{T-1}$ and there exists a set of signals $[w_b, \bar{w}]$ that would increase $\hat{\rho}_{T-1}^B$ enough that B would choose Y in period T regardless of the trial outcome. Once again, In this case, the argument in proposition 1 shows that there is no reporting strategy φ^{T-1} in which the seller is able to credibly convey any of these signals.

This result leads to the final proposition of the paper.

Proposition 3. *There exists no equilibrium reporting rule φ^t in any period $t=0, \dots, T$ such that a_t is a non-trivial function of the period t message m .*

Proof. Given that no reporting strategy of S can affect the actions of B in periods T and $T-1$ we can use an analogous argument to that of proposition 2 to show that this is also true in period $T-2$ and so forth.

3. Conclusions

The reasoning behind the main result of this paper is quite simple. Given that the seller always has incentive to misrepresent bad news, he can never credibly convey good news to the buyer. As such the seller can never commit to a reporting strategy which would give him a strictly higher expected payoff than transmitting no information at all. On the buyer's side, this means that the sequence of actions of the buyer is always a trivial function of the messages sent by the seller.

This negative result points to several directions for future work. One way in which the model could be extended would be to allow for multiple sellers. As economists we are well aware of the fact that competition forces often have a dramatic effect on individual behaviour. However, it is unclear if competition between multiple sellers would have any affect in the model presented here as it stands. The reason for this is that each seller would start the game with the same information as the buyer and thus would face a similar credibility problem as the one that arises in the model with one individual seller. A potential way to circumvent

²For instance, see Berry & Fristedt [8] and Kakigi [9].

this problem would be to make the (perhaps more realistic) assumption that each seller knows the type of his asset at the outset of the game. This would allow for the possibility of a separating equilibrium in which sellers of good assets would be willing to transmit bad information while sellers of “bad” assets would only want to transmit good information.

A perhaps critical assumption in the model is that the signal observed by the seller and the trial outcome of the asset are independent in each period conditional on y . Relaxing this assumption would make the model more closely aligned with models of reputational cheap talk and insiders. In this case, the buyer would be able to better evaluate the seller’s message based on the trial outcome. As such, the seller would face a tradeoff between maintaining a reputation for credibly transmitting short term information and maximizing the expected amount of experimentation of the buyer. An interesting consequence of relaxing this assumption is that a cutoff belief strategy may not be optimal for the buyer, in the sense that if information about the likelihood of success of the asset can be transmitted in a period then there may be an optimal switching policy which gives a higher expected payoff than playing even the best possible arm indefinitely.

REFERENCES

- [1] M. Rothschild, “A Two-Armed Bandit Theory of Market Pricing,” *Journal of Economic Theory*, Vol. 9, No. 2, 1974, pp. 185-202. [doi:10.1016/0022-0531\(74\)90066-0](https://doi.org/10.1016/0022-0531(74)90066-0)
- [2] D. Bergemann and J. Välimäki, “Learning and Strategic Pricing,” *Econometrica*, Vol. 64, No. 5, 1996, pp. 1125-1149. [doi:10.2307/2171959](https://doi.org/10.2307/2171959)
- [3] H. Bar-Isaac, “Reputation and Survival: Learning in a Dynamic Signalling Model,” *Review of Economic Studies*, Vol. 70, No. 2, 2003, pp. 231-251. [doi:10.1111/1467-937X.00243](https://doi.org/10.1111/1467-937X.00243)
- [4] V. P. Crawford and J. Sobel, “Strategic Information Transmission,” *Econometrica*, Vol. 50, No. 6, 1982, pp. 1431-1451. [doi:10.2307/1913390](https://doi.org/10.2307/1913390)
- [5] R. J. Aumann and S. Hart, “Long Cheap Talk,” *Econometrica*, Vol. 71, No. 6, 2003, pp. 1619-1660. [doi:10.1111/1468-0262.00465](https://doi.org/10.1111/1468-0262.00465)
- [6] M. Ottaviani and P. N. Sørensen, “Reputational Cheap Talk,” *RAND Journal of Economics*, Vol. 37, No. 1, 2006, pp. 155-175. [doi:10.1111/j.1756-2171.2006.tb00010.x](https://doi.org/10.1111/j.1756-2171.2006.tb00010.x)
- [7] R. Benabou and G. Laroque, “Using Privileged Information to Manipulate Markets: Insiders, Gurus and Credibility,” *Quarterly Journal of Economics*, Vol. 107, No. 3, 1992, pp. 921-958. [doi:10.2307/2118369](https://doi.org/10.2307/2118369)
- [8] D. A. Berry and B. Fristedt, “Bernoulli One-Armed Bandits—Arbitrary Discount Sequences,” *Annals of Statistics*, Vol. 7, No. 5, 1979, pp. 1086-1105. [doi:10.1214/aos/1176344792](https://doi.org/10.1214/aos/1176344792)
- [9] R. Kakigi, “A Note on Discounted Future Two-Armed Bandits,” *Annals of Statistics*, Vol. 11, No. 2, 1983, pp. 707-711. [doi:10.1214/aos/1176346176](https://doi.org/10.1214/aos/1176346176)