Experimental Results on Sponsored Search Auction:
Comparison of GSP and VCG

Emiko Fukuda, Yukihiko Funaki, Yoshio Kamijo, Michiharu Masui and Ai Takeuchi

April, 2009

Abstract

Sponsored links are a kind of Internet ads and its market has grown rapidly in recent years. This paper compares, using laboratory experiment, the performance of two auction rules for sponsored links; Vickrey-Clarke-Groves mechanism (VCG), which is an incentive compatible mechanism with dominant strategy equilibrium; and “generalized second price” auction (GSP), which is commonly used by the major search engines. In the experiment, we observed no difference in the distribution of the revenues for a search engine in VCG and GSP, and equilibrium outcomes resulted more frequently in VCG. Furthermore, outcomes in VCG were more efficient than the outcomes of GSP.

Key Words: Sponsored Search Auctions, Generalized Second Price Auctions, Vickrey-Clarke-Groves Mechanism, Laboratory Experiments

JEL Codes: C72, C91, D44

1 Introduction

The market for internet advertisements, which is said to first have appeared in 1994, is increasing rapidly. For instance, it was reported by Interactive Advertising Bureau\(^1\) that Internet ad spending in USA was more than 21 billion dollars in 2007. Also in Japan, Dentsu\(^2\) reported that Internet ad spending

\(^{*}\)Very Preliminary. Please do not quote.

\(^{†}\)This experiment was supported by the Grant-in-Aid for the Global COE “Political Economy of Institutional Construction” from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

\(^{1}\)http://www.iab.net/

\(^{2}\)http://www.dentsu.co.jp/marketing/ (in Japanese)
in 2008 was 698.3 billion yen. Sponsored links are a kind of internet ads that are displayed with the results of a search using a search engine. It is also called Pay for Performance advertisements (P4P ads), since the advertiser pays the rate to the search engine only when an user actually clicks the ad. According to Interactive Advertising Bureau and Dentsu, search revenues mark up 41% of total internet ad spending in USA and 23% of the total in Japan The market of sponsored links is now attracting attention as a big field of business, and its stability might be a consideration for both advertisers and search engines.

In this market, each advertiser places an ad for a keyword or several keywords. When a search engine user clicks the ad, the advertiser achieves ad effect for 1-click and pays the rate to the search engine. An advertisement’s total ad effect and payment per period is determined by how many clicks it achieves, and the number of clicks per period depends on the position of the ad. Ads placed in higher positions are clicked more frequently, so, the higher positions can be traded at higher prices. Search engines naturally employed an auction system to allocate ad spots.

“Generalized second price” auction (GSP), and or the modified version of it is the most widely used auction mechanism for selling sponsored links. Until 2007, Overture (Yahoo!) employed GSP, which allocates ad spots to the advertisers according to the descending order of the bids, i.e. the top position is allocated to the bidder with the highest bid, the second position is to the bidder with the second highest bid, and so on. Then each bidder pays the next lower bid per click, every time a search engine user clicks the bidder’s link. Presently, both Overture (Yahoo!) and Adwords (Google) uses a modified GSP. In the modified GSP, ad spots are allocated in the decreasing order of the ranking scores, which are determined for each advertiser \(i\) as the product of \(i\)’s bid and estimated click-through rate (CTR) \(w_i\) of \(i\) (see Aggarwal et al. (2006) for detail.) The modified GSP is like a weighted GSP, where CTRs are taken into account as the weights to calculate the expected revenue. In other words, the original GSP can be considered as a case where all advertisers have identical CTRs. Since the modified GSP are based on “generalized second price” auction, GSP can said to be the most widely used auction system in the current world.

GSP is similar to the Vickrey-Clarke-Groves mechanism (henceforth, VCG); when there is only one ad spot, GSP is exactly the same as VCG. However, while the VCG is an incentive compatible mechanism for selling multiple objects, GSP with more than one objects does not have a dominant strategy equilibrium and truth-telling is not an equilibrium. In Edelman and Ostrovsky (2007) they
reported that bids and revenues of GSP might be unstable, and its instability caused by bidders’ strategic behavior may decrease revenues for search engines seriously. In contrast, the bids and revenues of VCG can expected to be stable. Edelman and Ostrovsky (2007) concluded that switching from GSP to a new mechanism would be risky and costly for the search engines, but VCG might be an alternative auction system to reduce the problem of strategic behavior.\(^3\)

On the other hand, Edelman et al. (2007) newly defined “locally envy-free” equilibria of GSP in which bidders’ strategies are restricted, then showed that the revenue in the equilibria are greater than or equal to the revenue in the dominant strategy equilibrium of VCG (see Section 2 for details). That is, if a locally envy-free equilibrium and the dominant strategy equilibrium are achieved in GSP and VCG, respectively, GSP is more profitable than VCG. Varian (2007) also studied properties and relations among a locally envy-free equilibrium, the symmetric Nash equilibrium and the Nash equilibrium in GSP. He also observed that the Nash equilibria of GSP described the actual transaction prices of ad spots. However, since each bidders’ evaluation are not informed in the market, it is uncertain that the equilibrium are really achieved.

In this paper, we use a laboratory experiment to compare the performance of the two auction mechanisms, VCG and GSP. The theories suggest that the revenues for search engines will be higher in GSP, under the assumption of the common knowledge of evaluation of all advertisers. Although it is reasonable that others’ bids and CTRs of each ad spots become common knowledge through repetition, whether the evaluation would also become a common knowledge is questionable. Therefore, investigating the performance of the two auction mechanisms in experiments where the evaluations are not common knowledge would provide an additional ground for the evaluation of the two mechanisms. Also, the use of laboratory experiments allows for analysis on outcomes and equilibrium. Without the use of laboratory experiments where it is possible to know the evaluation of the ad spots, it is not possible to check whether the outcomes are in equilibrium or not. This experiment additionally allows us to compare the stability, revenues, efficiency and individual bidding behavior.

This paper is organized as follows. In section 2, we review the definitions and properties of GSP auction and VCG mechanism, and provide theoretical predictions based on Edelman and Ostrovsky

\(^3\)Incentive compatibility is usually focused in the auction theory as a desirable property. Aggarwal et al. (2006a) presented an incentive compatible auction, called laddered auction, taking into account the weights that are used in Adwords’ weighted GSP. Also in Aggarwal et al. (2006b) new auctions were proposed in which the advertisers not only submitted a bid but their preference for positions of ad spots. In particular they studied so-called prefix position auctions, and proposed the top-down auction mechanism, which is an incentive compatible prefix position auction.
(2007) and Edelman et al. (2007). In section 3, we state our hypothesis. Section 4 describes the experimental procedure and the results of the experiment are reported in section 5. We end with some concluding remarks in section 6.

2 Theoretical Predictions

2.1 Auctions

This section is devoted to the definitions of auctions for selling multiple ad spots. As common notation, we denote the set of advertisers (bidders) by $N = \{1, 2, \ldots, n\}$, the set of ad spots by $\{1, 2, \ldots, K\}$. In real life sponsored search auction, we can generally observe that the number of ad spots are determined to be same as $n$, so we assume $K = n$ in the following. For each ad spot, we can estimate the probability of being clicked, so called “click-through rate” or CTR. Let $\alpha_k$ be the expected number of clicks per period of the $k$-th highest ad spot\footnote{Although CTRs actually depend not only on positions of ad spots but also on adopted auction system and keywords, etc., here we assume that CTRs are identical among advertisers and that $\alpha_k$ depends only on the position $k$ where the ad is allocated: $\alpha_1 > \alpha_2 > \cdots > \alpha_K$.}. Each advertiser $i$ now estimates his/her ad effectiveness per click, call it “evaluation” or “value” and denote by $x_i$. If bidder $i$ obtains $k$-th highest position of the ad spots then the expected total ad effectiveness per period can be denoted by $\alpha_k x_i$. We denote bidder $i$’s bid by $b_i$ and bids by $b = (b_1, \ldots, b_n)$\footnote{For each keyword, we assume that bid is closed just before a user search for the word.}.

In 1997 when sponsored search auction was introduced, “generalized first price” auction was first used in USA. However, since bidders can change their bids frequently in the auction, transaction prices of ad spots were uncertain and unstable. Edelman and Ostrovsky (2007) pointed out, from the bidding data between June 2002 and June 2003, that bids tended to form a typical “sawtooth” pattern when two or more bidder existed in Overture’s generalized first price auction. This sawtooth pattern was caused by advertisers’ strategic behaviors to win a bid for a high position at a lower price, and reduced revenue for search engines as well as market efficiency. To avoid the inefficiency of a generalized first price auction, Adwords introduced GSP from February 2002, and after recognizing its advantages Overture also switched to GSP.

**Generalized Second Price (GSP) auction:** Given bids $b = (b_1, b_2, \ldots, b_n)$, GSP allocates ad spots and determine the payment and the utility for each bidder in the following manner:
**Ad spots allocation:** Ad spots are allocated in the descending order of the bids $b = (b_1, b_2, ..., b_n)$; for each $k \leq n$, let the bidder bidding the $k$-th highest bid be $d(k)$, then $k$-th ad spot ($k \leq K$) is allocated to $d(k) \in N$.

**Payment:** If bidder $i$ obtains $k$-th highest position (i.e. $i = d(k)$), he/she pays $b_{d(k+1)}$ per click to a search engine. Thus his/her total payment is $p_i(b) = \alpha_k b_{d(k+1)}$ where $b_{d(n+1)} = 0$.

**Utility (Payoff):** If $i = d(k)$ and bidder $i$’s value is $x_i$ then his/her utility is $u_i(b) = \alpha_k x_i - p_i(b)$.

The GSP succeeded to reduce the notable instability of the bids that was observed in the generalized first price auction. Edelman and Ostrovsky (2007) however observed there were still strategic biddings in GSP and bids were unstable. They suggested to introduce VCG to stabilize transaction prices and revenue for the search engines.

**Vickrey-Clarke-Groves (VCG) mechanism:** Given bids $b = (b_1, b_2, ..., b_n)$, VCG allocates ad spots and determine the payment and the utility for each bidder in the following manner:

**Ad spots allocation:** VCG allocates ad spots in the same manner as GSP (i.e. in the descending order of $b$).

**Payment:** Bidder $i$’s payment in VCG is determined to be the negative externality that $i$ imposes on the other bidders. Therefore, if bidder $i$ obtains $k$-th highest ad spot (i.e. $i = d(k)$), his/her payment is as follows.

$$p^V_i(b) = \left[ \sum_{j=1}^{k-1} \alpha_j b_{d(j)} + \sum_{j=k}^{n} \alpha_j b_{d(j+1)} \right] - \left[ \sum_{j=1}^{k-1} \alpha_j b_{d(j)} + \sum_{j=k+1}^{n} \alpha_j b_{d(j)} \right]$$

$$= \sum_{j=k}^{n} \alpha_j b_{d(j+1)} - \sum_{j=k}^{n-1} \alpha_{j+1} b_{d(j+1)},$$

where $k = n$, $p^V_{d(n)} = 0$. $p^V_i(b)$ can be represented as a recursive function; $p^V_i(b) = (\alpha_k - \alpha_{k+1}) b_{d(k+1)} + p^V_{d(k+1)}(b)$.

**Utility (Payoff):** If $i = d(k)$ then his/her utility is $u^V_i(b) = \alpha_k x_i - p^V_i(b)$.

VCG is commonly known to be an incentive compatible mechanism, and, in general, it is exactly same as GSP if there is only one ad spot.
2.2 Review of Theoretical Results

Since bidders can change their bids in the market of sponsored links, sponsored search auctions can be modeled as infinitely repeated games. In this game we can assume each bidder has private information of his/her own value $x_i$ and gradually learn the values of others. There are, however, numerous equilibria in an infinite repeated game, and bidders are required to use extremely complex strategies to support such equilibria. Besides, usually advertisers bid for more than one keyword.

Edelman et al. (2007) therefore studied convergence points of the bidding process with the following assumptions:

(S1) As a result of bidders’ learning, all evaluations are common knowledge.

(S2) In convergence points (i.e., stable bids) $(b_i)_{i \in N}$, each bid $b_i$ is a best response to the others $b_{-i} = (b_1, ..., b_{i-1}, b_{i+1}, ..., b_n)$. That is, the stable bids are an equilibrium of static game with complete information where (S1) holds.

Formally, let us define a normal form game with complete information induced by GSP by $\Gamma = (N, [0, \infty)^n, (u_i)_{i \in N})$ where $[0, \infty)$ is strategy set for each bidder $i \in N$. In the same manner, we also define a game with complete information induced by VCG by $\Gamma^V = (N, [0, \infty)^n, (u_i^V)_{i \in N})$.

Here we note two remarks about GSP and VCG in Edelman et al. (2007).

**Remark 2.1.** For GSP and VCG, the following holds:

1. For each bidder in VCG, the truth-telling is his/her dominant strategy. Thus $\Gamma^V$ has the dominant strategy equilibrium.

2. If every bidder $i$ bids the same $b_i$ under both of GSP and VCG, then $p_i(b) \geq p_i^V(b)$ for any $i \in N$.

While VCG has the truth-telling dominant strategy equilibrium, there is no dominant strategy equilibrium in GSP. So we use Nash equilibrium to analyze GSP.

**Definition 2.1.** $b = (b_1, b_2, ..., b_n)$ is a Nash equilibrium of $\Gamma$ if for all $i \in N$

\[
\begin{align*}
    u_i(b) &= \alpha_k x_i - p_i(b) = \alpha_k (x_i - b_{d(k+1)}) \geq \alpha_l (x_i - b_{d(l)}) & \text{for all } l < k \\
    u_i(b) &= \alpha_k x_i - p_i(b) = \alpha_k (x_i - b_{d(k+1)}) \geq \alpha_l (x_i - b_{d(l+1)}) & \text{for all } l > k
\end{align*}
\]  

(1)

where $d(k) = i$. 6
For a game induced by GSP $\Gamma$, there are numerous Nash equilibria. Edelman et al. (2007) thus focused on simple strategies and supposed that (S3) each bidder uses strategies which force out the bidder who occupies the position immediately above.

Under the assumptions (S1) – (S3), Edelman et al. (2007) call the bids “locally envy-free” if, in the state, every bidder has no incentive to change his/her bid to force out the bidder immediately above. And they introduced “locally envy-free” equilibrium for GSP.

**Definition 2.2.** \(b_i\) is locally envy-free at profile \(b = (b_1, b_2, ..., b_n)\) if \(u_i(b) = \alpha_k x_i - p_i(b) \geq \alpha_{k-1} x_i - p_{d(k-1)}(b)\), where \(d(k) = i\). Profile \(b\) is locally envy-free if for each \(i\), \(b_i\) is locally envy-free at \(b\). Finally, profile \(b\) is a locally envy-free equilibrium of the normal form game \(\Gamma\) induced by GSP if \(b\) is an equilibrium and is locally envy-free.

\(\Gamma\) may have multiple locally envy-free equilibria, Edelman et al. (2007) constructed a certain locally envy-free equilibrium \(b^* = (b^*_1, b^*_2, ..., b^*_n)\) as follows: For any \(k = 2, ..., K\), let \(b^*_{d(k)} = p^V_{d(k-1)}(\bar{b})/\alpha_{k-1}\) where \(\bar{b}\) be the dominant strategy equilibrium of a game induced by VCG $\Gamma^V$. For \(k = 1\), \(b^*_1 = b^*_{d(1)} = x_{d(1)}\).

**Theorem 2.1** (Edelman et al. (2007)). *Strategy profile \(b^*\) is a locally envy-free equilibrium of \(\Gamma\). Furthermore (i) \(p_i(b^*) = p^V_i(\bar{b})\) for all \(i \in N\), and (ii) \(\sum_{i \in N} p_i(b') \geq \sum_{i \in N} p_i(b^*) = \sum_{i \in N} p^V_i(\bar{b})\), where \(b'\) are all locally envy-free equilibrium strategy profile.*

The theory suggests that if a locally envy-free equilibrium and the dominant strategy equilibrium are achieved in GSP and VCG, respectively, GSP brings more benefit to search engines than VCG. However, it is uncertain that the locally envy-free equilibrium is archived in the real market and players execute the dominant strategies in VCG when VCG is used instead of GSP. Furthermore, this uncertainty has search engines hesitate to switch their auction systems, or introduce VCG, although Edelman and Ostrovsky (2007) suggested search engines could stabilize the bids and the revenue by introducing VCG since the bids in GSP might be unstable.

Varian (2007) in addition studied relations among the locally envy-free equilibrium, Nash equilibrium and symmetric Nash equilibrium (SNE).

**Definition 2.3.** \(b = (b_1, b_2, ..., b_n)\) is a symmetric Nash equilibrium of the normal form game \(\Gamma\) induced by GSP if no bidder can improve his/her utility by exchanging bids with any other bidders. Formally,
for any $k \leq K = n$,

$$u_i(b) = \alpha_k x_i - p_i(b) \geq \alpha_l x_i - p_l(b)$$

(2)

for all $l \neq k$, where $d(k) = i$.

Varian (2007) studied properties of the locally envy-free equilibrium, the symmetric Nash equilibrium and the Nash equilibrium in GSP, and, for example, noted that if bids $b$ is SNE it is Nash equilibrium. Clearly, any symmetric Nash equilibrium is a locally envy-free equilibrium, because $b$ is locally envy-free if formula (2) in Definition 2.2 holds when $l = k - 1$. Varian (2007) also provided an empirical observation that the Nash equilibria of GSP described the basic properties of the transaction prices of ad spots, from the Adwords’ data. Here again, however, we do not know if the equilibrium is actually achieved.

3 Hypothesis

As reported in the previous section, Edelman et al. (2007) supposed assumptions (S1)–(S3) and studied the sponsored search auction by modelling it as a static game with complete information, since it is difficult to analyze by dynamic models. Varian (2007) and other theoretical literatures also supposed the full information (S1). Indeed, each bidders can easily observe the expected number of clicks of each ad spot by Adwords’ reports or through third party companies so-called “Search Engine Managers.” It may, however, be difficult for them to learn other bidders’ evaluation through the bidding process, and the actual sponsored search auction may be a game of incomplete information. In order to compare the performance of the two auction mechanisms under a situation where the value of other bidders are unknown, we’ve conducted a laboratory experiment.

In the experiment, subjects will be new to the situation, so one can interpret it as exemplifying a new market which has just been generated. If the value of other bidders, or an equilibrium bids are learned through repetition, we can expect the subjects’ bid to stabilize at some level and that the ad spots would

---

6Varian (2007) dealt with a non-weighted GSP while he used Adwords’ data.

7Using laboratory experiment in this auction has merits over the econometric analysis using the actual data or a field experiment. First, it makes comparison of individual behavior under VCG and GSP possible, which is not possible just from the existing sponsored search auction data. Second, even if one were able to conduct an extremely expensive field experiment— asking for a change in auction rule for month or two to attain data— laboratory experiment still has the advantage that the values are known to the experimenter. Without the information of the values, one cannot analyze the efficiency or whether the outcomes are in equilibrium or not. Therefore, using laboratory experiment is an appropriate method to compare the performance of the two mechanisms.
be allocated according to the order of the values, improving the efficiency.

(H1) Subjects’ bid become stable at some level in both VCG and GSP.

(H2) Efficiency improves with repetition in both VCG and GSP.

Since the ad spots will be allocated according to the order of the values in the equilibrium of both mechanisms (dominant strategy equilibrium in VCG and locally envy-free equilibrium in GSP), theory expects that outcomes will be efficient in both mechanisms. However, when the evaluation is not common knowledge, whether the efficient outcome will be achieved or not may not be as clear as the theory predicts. In VCG, the information about others’ values are unnecessary since there is a dominant strategy equilibrium, and because its unique, there is no problem of the coordination of the equilibrium. This might make it easier for the ad spots to be allocated according to the order of the values in VCG.

(H3) Efficiency achieved in VCG is higher than that achieved in GSP.

In order to check whether the three equilibrium is achieved in GSP, we can compare the relative frequency of Nash equilibrium, locally envy-free equilibrium, and symmetric Nash equilibrium outcomes in both mechanisms.

(H4) There is no difference in the relative frequency of outcomes in Nash equilibrium, locally envy-free equilibrium, and symmetric Nash equilibrium in VCG and GSP.

Finally, we examine the following hypothesis to check if the result of the theorem of Edelman et al. (2007) would hold under the condition where the game is not full information.

(H5) The revenue for search engine is higher in GSP than VCG.

4 Experiment Design

In order to test for these hypotheses, we’ve conducted an experiment with two treatments, VCG and GSP. First, we introduce the parameters used in the experiment and how the values are determined in the experiment. Then describe the procedures of the experiment.
4.1 Parameters in the Experiment and Determination of Values

In both treatments, all the parameters were set the same, so the only difference between VCG and GSP is how payments are calculated. The number of players in a group, \( n \), and the number of ad spots, \( K \), were both set to 5, so that all players could gain an ad spot. CTR for each ad spots were set to \((\alpha_k)_{k=1}^{5} = (100, 75, 42, 33, 22)\) and were constant across the whole experiment.\(^8\) Subjects’ bids were restricted to \([0, 10000]\) for programming purpose.

The value of each click were set to \((x_i)_{i=1}^{5} = (120, 100, 80, 60, 40)\) and were kept constant across the whole treatment. The subjects were informed in the instruction that the computer determines the value automatically; their value might be different from the value of four other group members; and that they would not be able to know the other subjects’ value, but the difference between the highest and the lowest value is at most 100. Three main reasons, other than it being simple and avoids misunderstanding by the subjects, have led us to use this experiment design for the values. First, we kept the value constant across the whole treatment, because the actual sponsored search auctions have such a tendency. Unlike the usual auction like Christie’s auction where different object is sold every period, sponsored search auctions sells the same ad spots. Therefore, it is likely that the participants have the same value across periods. Secondly, this method avoids giving the subjects the exact range of the distribution of the value, and prevents them from having a sense of the relative position of their value among the group members. Finally, although the question of the effect of difference in the value on the subjects’ behavior is an interesting problem on its own, our interest in this first experiment is not on this effect of difference in the distributions of values, but on the difference of the individuals behavior induced by the two different mechanisms. Therefore, keeping the value constant across the treatment and groups allows for more meaningful comparison of the two treatments.

4.2 Experiment Procedure

The experiment was conducted on November 2008 at Waseda University Political Economics Experimental Room. Subjects were undergraduate school students of Waseda University whom were recruited through a university’s web page for part time jobs. Their majors were not restricted to economics but diverse, and we recruited about the same number of subjects from each department.

---

\(^8\)CTRs were chosen arbitrarily. We had fixed it throughout the experiment to simplify the experiment and to avoid misunderstanding by the subjects.
Experiment consisted of four sessions, two for each treatment. Each session had fifteen subjects, and each subject were only allowed to participate in one session. We had incorporated the between-subject design, so each subject participate in one treatment only. We conducted two sessions per day, and to control for the possible, although unlikely, effect of the morning and afternoon sessions and the effect of the day, we conducted VCG in the morning session and GSP in the afternoon session on one day, and switched the order in the other day.

Each session proceeded in the following order. First, subjects were seated anonymously to the computer terminal. They were handed a recording sheet, to record the results of each period, and a summary of the instructions. All the experimental materials are available from authors among request. After all the participants had taken their seats, the experimenter told the subjects to read the instruction slides shown on the computer screen on their own, at their own pace. After everyone had finished reading the instruction slides, understanding test was handed out, which was, upon completion, checked one by one by the experiment staffs. After every subject had completed the test and had correct answers, answers were reviewed once again using the over-head projector. This was done to inform the subjects that all participants had taken the same instruction and the understanding test. After explaining the answers to the understanding test, two practice rounds were conducted to have the subjects get the feel of the computer program, which was programmed and conducted with the software z-Tree (Fischbacher 2007). In the practice rounds, every subject had a value of 100, and we asked them to bid a certain number so as to control for the possible learning from the practice rounds.

During the actual treatment, subjects participated in a sponsored search auction using either GSP or VCG rule, which only differs in the determination of payments. Subjects participate in one of the two auctions for 40 rounds in total. The 40 rounds were divided into two subgroups of 20 rounds, which we sometimes refer to as first and second half. The value and the group remained constant within each subgroup, but once after the twentieth round was finished, the value and the groups were reshuffled.

9By using the power point slide instructions and allowing the subjects to read the instructions at their own pace, we believe that subjects better understood the rules of the auctions. Since the rule of the auction was difficult, this use of power point slides were necessary to insure their understanding, even at some cost of public knowledge of the instructions. We have compensated for the public knowledge of the instruction by reviewing and explaining the answers to the understanding test out loud.

We checked whether we had succeeded in explaining the auction rule to the subject, and whether subjects understood the instruction by asking two questions in the after experiment questionnaire. In the direct question, we asked whether “The instruction was easy to understand” using 5 point scale from “Strongly Disagree (= 1)” to “Strongly Agree (= 5).” The mode was “Agree” and average points were 3.78. Also, in more indirect question, we asked when they were choosing their bid, whether they “Decided randomly”. This was asked using 7 point scale from “Strongly Disagree (= 1)” to “Strongly Agree (= 7).” The average point was 1.88 and the mode answer was “Strongly Disagree.”
in the following manner. We shuffled the value so that the subjects with highest value in the first half would receive the lowest value in the second half, and the subjects with the second highest in the first half would receive the second lowest in the second, and so on. The groups were randomly matched in a way that this reshuffling of the values was possible. This design not only allows us to analyze a situation where the subjects with the same value repeatedly participate in an auction, but at the same time, reduce the possible difference in the final earning caused by the difference in the randomly assigned value of the participants.

The computer screen for each round proceeded as follows. First, subjects decide on their bid. On the next screen, they enter whether they had, in deciding their bids, aimed for certain position, and if so to which position of the ad spots. We emphasized to the subjects that this question was just a questionnaire and was unrelated with their payment. This allows us to analyze the relationship between the subjects’ bids and their aim. Then, after every subject had enter their aim, the result screen was shown to the subjects, where they received feedback information on: all the group members’ bid and payment (in the descending order of bids), his/her own bid, the earned position, revenue, payment, and payoff. After the 20th and 40th round, they received additional information on the total payoff s/he earned in the first and the second half.

After the experiment, participants answered a short questionnaire while the experimenter and the staffs prepared for the payment. The subjects were paid in cash for the participation fee and the sum of points they had earned in the forty rounds. The participation fee differed from 800 Yen to 1100 Yen depending on the length of the experiment, and the exchange rate of the points was 100 points = 1 Yen. The average payment was about 1990 Yen, and the experiment lasted for about two to two and a half hours.

5 Experimental Results

The main motivation of this experiment is to compare the performance of the two auctions. In addition, we examined the bidding behavior of individuals to have a better understanding of the two mechanisms. In this section, we first analyzes the individual bidding behavior, including whether the bids stabilize at some level (H1), then, compare the number of equilibrium outcomes (H4), revenues (H5) and efficiency (H2, H3) of VCG and GSP.
In the analysis, we sometimes refer to the true ranking of individual \(i\), which we denote as \(T_i = \#\{x_j | x_j \geq x_i, j \in N\}\) (e.g., in this experiment, subjects with \(x_i = 80\) have \(T_i = 3\)).

### 5.1 Individual’s Bidding Behavior

In this subsection, we will treat each individuals’ data in each period as an independent observation. Each treatment had two sessions with 15 subjects each, and each subject participated in 40 rounds of decision making, resulting in 1200 observations for each treatment. However, we most of the times treat subjects with different values separately. In such cases, the number of observations for each value are the same as that of number of groups, i.e., 12 observations for each value for each round.

<table>
<thead>
<tr>
<th>Value</th>
<th>VCG</th>
<th>GSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>164.95 (641.87)</td>
<td>93.61 (154.07)</td>
</tr>
<tr>
<td>100</td>
<td>106.63 (54.26)</td>
<td>56.61 (19.88)</td>
</tr>
<tr>
<td>80</td>
<td>79.45 (18.97)</td>
<td>60.01 (20.70)</td>
</tr>
<tr>
<td>60</td>
<td>68.14 (35.22)</td>
<td>29.85 (16.49)</td>
</tr>
<tr>
<td>40</td>
<td>55.86 (46.51)</td>
<td>22.87 (17.19)</td>
</tr>
</tbody>
</table>

Table 1: Average (and Standard Deviation in Parentheses) of Bids for each Value

![VCG and GSP Figure](image)

Figure 1: Comparison of Distribution of Bids in VCG and GSP

As we reviewed in Remark 2.1, if bidders bid the same amount under the two mechanisms, the payment to the search engine would be higher in GSP than in VCG. Table 1 lists the average and standard deviation (in parentheses) of bids for each value, and Figure 1 shows the scatter plot of bids and the average for each true ranking. From these table and figure, one can observe that the average bids are lower in GSP than in VCG, for all values. The distribution of bids were significantly different among VCG and GSP at 1% for all values using Wilcoxon rank sum test. Looking within each treatment, it is clear that the average of bids for each of the value is different, and it tends to be lower for lower values. For all pairwise comparisons, the difference in the distribution of bids for each value were significant at 1% level for VCG, and all but the difference between the distribution of bids for \(T_i = 2\) and \(T_i = 3\) were significant at 1% level for GSP (pairwise Wilcoxon rank sum test using p-adjustment method of holms). Although the average bids in VCG looks close to the value compared to GSP, we can still reject
the null hypothesis that the true location of bids equals to the value in VCG for $T_i = 1, 4,$ and $5$ at 1% level using the Wilcoxon signed rank test. While the bids in VCG looks close to the value, the bids in GSP looks close to $b^* = (120, \ 61.2, \ 48.3, \ 23.3, \ 13.3),$ the pair of bids which is locally envy-free equilibrium and achieves the same payoffs as dominant strategy equilibrium of VCG. When one compares the average bids to the $b^*$, we can reject the null hypothesis that true location is equal to $b^*$ for all values at 5% level (Wilcoxon signed rank test). The average bids are lower than $b^*$ for $T_i = 1$ and $2$ whereas higher for $T_i = 3, \ 4$ and $5$. Still, in GSP, all averages of bids are significantly lower than their value. To summarize, we have the following observation.

**O1** The average bids were lower in GSP than in VCG, and average bids are decreasing with the value in both treatments.

From the large standard deviation in Table 1, one may wonder whether the bids stabilized at some level or not. To analyze for such tendency, let us denote the absolute difference in bids of subject $i$ in period $t$ and $t + 1$ as $\Delta b_{i,t}$. Figure 2 shows the transition of the mean and the median of $\Delta b_{i,t}$. As one may infer from the high standard deviations of bids, subjects sometimes bid extremely large number. Since the mean of each period is affected by these extreme bids, we also plotted the median of $\Delta b_{i,t}$ for each period $t$.

The mean and the median in Figure 2 depicts that there are no difference in $\Delta b_{i,t}$ in the two treatments. Also, the bids seem to be more stable in the last 10 rounds than in the first 10 rounds. So, bids seem to have relatively stabilized at some level.

**O2** Transition of the median of $\Delta b_{i,t}$ were not different in VCG and GSP, and it tend to decrease over time in both treatments.

Figure 3 provides information about at which level the bids have stabilized. This graph shows, for VCG and GSP, the transition of average bids separately for each of the values. In VCG, the diamond points indicates the location of values $x_i$, whereas in GSP, they represent $b^*_i$. From this Figure, one can see that in VCG and GSP, the average final bids are lower in GSP than in VCG for each value. The subjects’ bids tend to stabilize at somewhere between his own value and one higher value in VCG and his $b^*$ and the one better $b^*$ in GSP (exceptions are average bids of subjects with $x_i = 120$ and $x_i = 100$). This kind of extreme over bidding is observed most frequently with subjects who has the highest value. This reminds us of the results of second price auction, where over bidding is frequently observed.
Figure 2: Mean and Median Absolute Difference of the Bid in Period $t$ and $t+1$

* Diamond points indicates the location for $x_i$ in VCG and $b_i^*$ in GSP.

Figure 3: Comparison of Transition of Average Bids Calculated Separately for Each Values
in GSP). This can be interpreted as a kind of spiteful behavior on the part of subjects. If one bids a little under the bid of a subject in one better position than him/herself, than s/he can increase the payment of that subject while keeping his/her payment constant. This is consistent with the answers to the after experiment questionnaire, where we asked whether s/he increased their bid in order to increase the payment of others in better positions.\footnote{In the questionnaire, we asked “Please tell us about how you decided your bid” and “I increased my bid so as to increase the payment of people in higher position” (translated from Japanese). For more analysis on other questions in the questionnaire, see Appendix.} They answered in 7 point scale from “totally disagree (= 1)” to “strongly agree (= 7).” Average point was 4.9 and most answered higher than “agree.”

One may wonder whether subjects were choosing bids so as to improve their utility compared to the previous period. We checked whether subjects’ choices were improving their utility, if all the other member made the same bids. Since it is an difficult task for the subjects to guess the bids of their group members, it is reasonable that they use the informations from the previous period in order to make their decisions. We say that subject $i$’s bid in period $t + 1$ is myopic improvement if $u_i(b_{i,t+1}, b_{-i,t}) - u_i(b_{i,t}, b_{-i,t}) \geq 0$, where $b_{i,t}$ is the bid of subject $i$ in period $t$ and $b_{-i,t}$ is the bid of group members other than $i$ in period $t$. Relative frequency of individual choices myopic improvement bids were 0.89 (1015/1140) for VCG and 0.90 (1023/1140) for GSP. Strictly improving bids were not so frequent in either treatment with 0.23 (266/1140) and 0.23 (259/1140) for VCG and GSP respectively. So there were no difference between VCG and GSP in the relative frequency of myopic improvement bids.

\textbf{O3} Relative frequency of myopic improvement bids were high and about the same in the two treatments.

\section*{5.2 Comparison of the Performance of VCG and GSP}

Now, let us compare the performance of VCG and GSP. In the rest of the analysis, we will look at the tendency of the group data. We will treat each group data in each period as an independent observation. There were 6 groups in each session (3 groups for first and second half) and two sessions per treatment, we thus have 12 groups for analysis. Each of these groups participated in 20 rounds of decision making. Therefore, in each treatment, the number of observations used for the analysis is 240.
5.2.1 Equilibrium

As we saw in Table 1, bids in VCG were different from the dominant strategy. As a result, all observed outcomes were not dominant strategy equilibrium. For comparison with the results of GSP, Nash equilibrium, locally envy-free equilibrium, and symmetric Nash equilibrium can also be defined for VCG, by exchanging the payment function $p(\cdot)$ with $p^V(\cdot)$ in definition 2.1, 2.2, and 2.3 respectively. Using these definitions, we compared the relative frequency of outcomes that were Nash equilibrium, locally envy-free equilibrium and symmetric Nash equilibrium in VCG and GSP.

![Figure 4: Relative Frequency of Locally Envy-Free Outcomes](image)

![Figure 5: Relative Frequency of Locally Envy-Free Equilibrium and Symmetric Nash Equilibrium Outcomes](image)

Figure 4 and Figure 5 compares the relative frequency of locally envy-free equilibrium outcomes and symmetric Nash equilibrium outcomes respectively in the two treatments. From these two figures and from some additional analysis, we have the following results concerning the outcomes and equilibrium. First, we can see that locally envy-free outcomes are not frequent in both treatments. Out of 240 outcomes in each treatment, the total number of outcomes which was locally envy-free were 28 for VCG and 7 for GSP. Although there were not many outcomes which were locally envy-free, there were more locally envy-free outcomes in VCG than in GSP. The difference in the distribution of group total of locally envy-free outcomes were significant at 10% level (Wilcoxon rank sum test, $W = 102.5$, p-value $= 0.06$)\(^{12}\). As for equilibrium outcomes, outcomes which were locally envy-free equilibrium and

\(^{12}\)We could not compute the exact p-value due to ties.
symmetric Nash equilibrium coincided in our experiment. Although very few, there were some of these equilibrium outcomes in VCG, but no outcomes were locally envy-free equilibrium in GSP (difference is significant at 1% level, using the Fisher exact test). In fact, there were only small number of Nash equilibrium outcomes in this experiment. The number of Nash equilibrium outcomes were 35 for VCG and 10 for GSP. So, in response to Hypothesis H4, we can summarize the observations regarding the equilibrium outcomes as follows:

O4. There were no outcomes in dominant strategy equilibrium for VCG. When one compare the outcomes in VCG and GSP, former has more outcomes in locally envy-free than the latter. Nash equilibrium, locally envy-free equilibrium nor the symmetric Nash equilibrium are much observed in either treatments.

5.2.2 Revenue of Search Engine

In the theorem of Edelman et al. (2007), it had been shown that the revenues for the search engine in locally envy-free equilibrium of GSP are greater than or equal to the revenue in dominant strategy equilibrium of VCG. We therefore expected that the average revenue for the search engine would be higher under GSP than in VCG.

Figure 6 shows the average revenues for the search engine and the average of the sum of revenues for the bidders for each treatment. First, by comparing the two left bars of this figure, it is clear that there

![Figure 6: Average Revenue of Search Engine and Sum of Payoff of the Bidders](image)

![Figure 7: Transition of the Average Revenue of Search Engine](image)
is no difference in the average revenue of the search engine in the two treatments. The average revenue of the search engine is 11616.14 for VCG and 11787.83 for GSP (standard deviations were 2302.7 and 2483.2 respectively), and this difference is not significant (t-test, t=-0.79, p-value=0.43). Also, there was no difference in the average sum of revenues obtained by the bidders in a group. The average was 13165.78 for VCG and 12710.76 for GSP (standard deviations were 2715.7 and 2505.7 respectively), and this difference is not significant (t-test, t=1.91, p-value=0.06). We also checked whether there is any tendency across periods. Since the groups were kept constant during the 20 rounds, it is possible that the subjects in each group learn to coordinate in the experiment, lowering the revenues for search engine. However, as one can see in Figure 7, there was no such downward tendency in GSP nor VCG. The revenue of the search engine was relatively stable across periods. This result, along with the results of the analysis on individual behavior, suggest that the reason which accounts for this indifference in the revenue raised in the two auction rules is not the unstableness of bids in GSP as expected by Edelman and Ostrovsky (2007), but the fact that the distribution of bids in GSP were close to $b^*$. 

5.2.3 Efficiency

Finally, although the theory expects that efficiency would be achieved in both mechanism, it may not be the case when the value of other bidders are unknown. As illustrated in the previous section, even the locally envy-free equilibrium is not reached in both VCG and GSP, so outcome might not be efficient either. Here, we use the number of subjects in a group who achieved the same rank as true ranking as a measure for efficiency, because the result of an auction is efficient if for all $i$, $i = d(T_i)$; every subject’s bid is in the order of their true ranking. Concerning efficiency, result of the experiment supported H3; more outcomes in VCG were efficient compared to GSP. The number of groups whose ad spots were allocated in the order of true ranking was 60 for VCG and 38 for GSP. This difference in the proportion of groups which achieved an efficient outcome was significant at 5% level (Fisher test, p-value = 0.017). We also analyzed the distribution of

13 Since our experiment includes repetitions, we calculated the mean for each group, and compared the distribution of group mean between VCG and GSP. The difference in the distribution of group mean of auctioneers’ revenue were insignificant, and so were the sum of utilities of the bidders (p-values of Wilcoxon rank sum test were 0.59 and 0.41 respectively.)

14 All the results presented in this section could also be confirmed using Kendall tau correlation, and group sum of absolute difference between true rank and earned position, which is $\sum_{i=1}^{5} |T_i - d^{-1}(i)|$. For more detailed analysis of efficiency using these measures, see appendix.
number of subjects in a group whose bid was in the order of their true ranking$(\#\{i|d(i) = d(T_i)\})$. The distributions were significantly different in the two treatments, with VCG having higher location than GSP (Wilcoxon rank sum test, p-value = 0.040). Also, in relation to H2, although efficient results were more frequently observed in VCG, in both treatments, the average number of subjects in a group whose bid was in the order of their true ranking was increasing as round proceeds. The results are shown in Figure 8. So, as is hypothesized in H2, efficiencies were improving in both treatments.

Efficient result were achieved more often in VCG than in GSP, and in both treatments, the efficiencies were improving with repetition.

6 Conclusion

This paper analyzed the results of a laboratory experiment conducted to compare two rules for sponsored search auctions, VCG and GSP. VCG has theoretically preferable characteristics: it is incentive compatible and has an unique dominant strategy equilibrium. GSP and the modification of it is the most widely used auction rule by the search engines. It is not either incentive compatible or has dominant strategy equilibrium, but may raise more revenues for the search engines. Edelman et al. (2007) defined locally envy-free equilibrium, and showed that all locally envy-free equilibriums in GSP raise more rev-
enue for search engines than the dominant strategy equilibrium in VCG. These theories on sponsored search auction studied it by modelling it as a static game with complete information. However, it may be difficult for the advertisers to figure out the value of other bidders even through repeated bidding process. To compare the two mechanisms under the condition of incomplete information on the evaluation of other bidders, we conducted a laboratory experiment.

The experimental results show that, unlike what was concerned in Edelman and Ostrovsky (2007), bids in GSP were stable as VCG (O2), and despite of the theorem by Edelman et al. (2007), the distribution of revenue for the search engine were not different in the two mechanisms (O5). Two reasons for these results are apparent in the data. First, dominant strategy equilibrium, locally envy-free equilibrium and symmetric Nash equilibrium were not frequently observed in either GSP or VCG (O4), and second, the bids in GSP were close to $b^*$, the pair of bids which give the same revenue to the auctioneer as the dominant strategy equilibrium of VCG. Therefore, in terms of stableness and amount of revenue, VCG and GSP yield similar results. However, when the evaluation of others’ values is unknown, one cannot simply assume that the outcome would be efficient. When one looks at the amount of locally envy-free outcomes and efficient outcomes, both were higher in VCG than in GSP (O4 and O6). Since efficiency is an important aspect for markets and mechanisms to fulfill, in this regard, experimental results suggest that VCG is a better auction rule than GSP.

For further research, it would be interesting to see if the findings of our paper would still hold when the distribution (mean and variance) of the values are changed. Also, although we mentioned in footnote 4 that in most cases, number of ad spots equal the number of participants of the auction $K = n$, testing to see whether the results would hold under the case where $K < n$ would be an interesting experiment.

References


