Online Ad Auctions: An Experiment [*]

Kevin McLaughlin[†]

Daniel Friedman[‡]

This Version: February 16, 2016

Abstract

A human subject laboratory experiment compares the real-time market performance of the two most popular auction formats for online ad space, Vickrey-Clarke-Groves (VCG) and Generalized Second Price (GSP). Theoretical predictions made in papers by Varian (2007) and Edelman, et al. (2007) seem to organize the data well overall. Efficiency under VCG exceeds that under GSP in nearly all treatments. The difference is economically significant in the more competitive parameter configurations and is statistically significant in most treatments. Revenue capture tends to be similar across auction formats in most treatments.

JEL-Classification: C92, D44, L11, L81, M3

Keywords: Laboratory Experiments, Auction, Online Auctions, Advertising.

[†]Economic Science Institute, Chapman University, CA 92866, kevin.mclaughlin5@gmail.com

[‡]Economics Department, University of California Santa Cruz, CA 95064, dan@ucsc.edu

1   Introduction

Online advertising has grown from almost nothing two decades ago into a major industry today. With worldwide revenue likely to exceed $180 billion in 2016 (Ember, 2015), it is the cash cow of several leading technology companies and is the lynchpin of the "new economy."
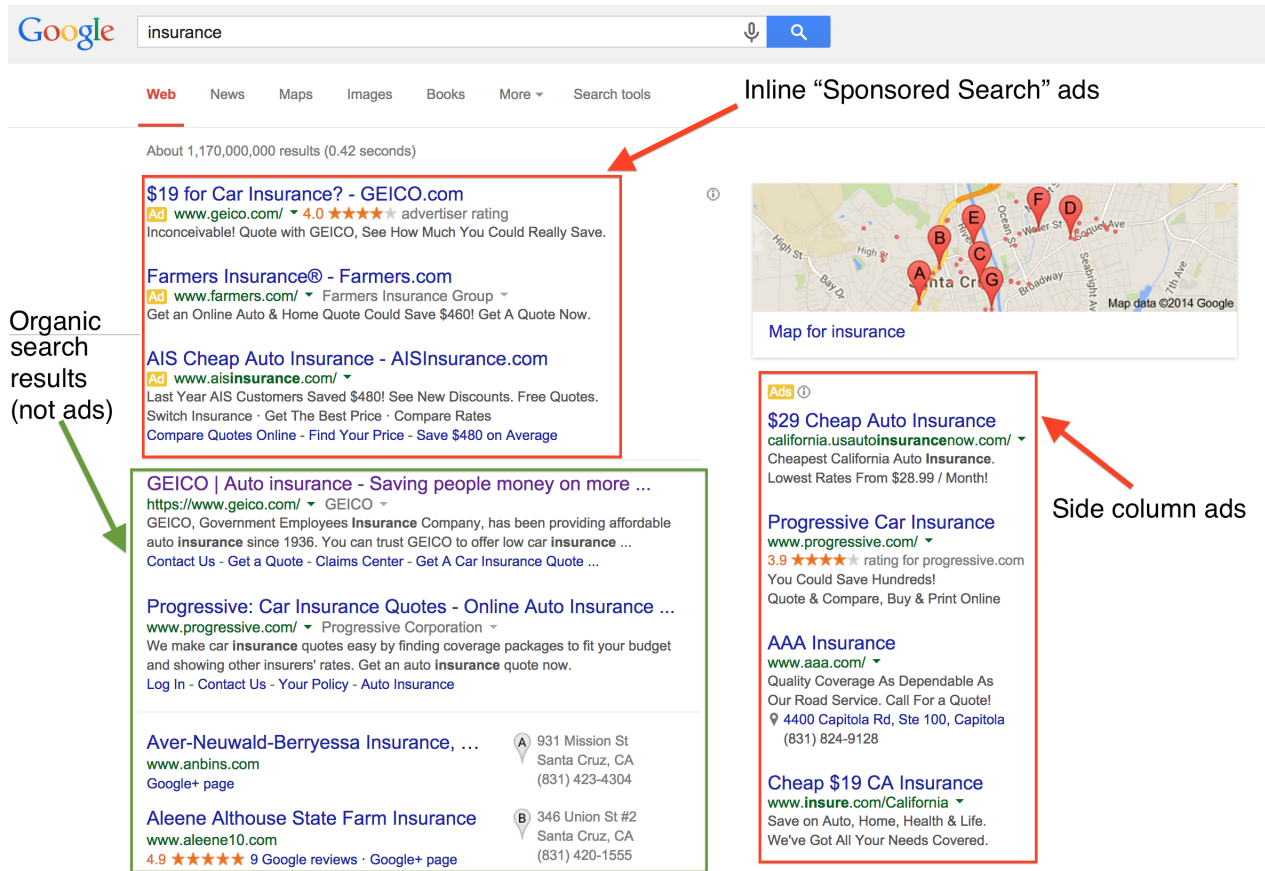

Figure 1: An Online Ad Display. An automated auction determines which ads go into the seven slots (in red boxes) in this Google search page.

Most online ad space is allocated via auctions. For example, Figure 1 shows a user's Google search page following the query "insurance." Besides the top-ranked links discovered by the search engine (shown in the green box), this user sees three "sponsored search" ads and four "side column" ads. An automated auction determines which ads go into those seven slots. Major search engines handle tens of thousands of queries every second, so the auctions for popular keywords are conducted in essentially continuous time.

Google adopted the generalized second price (GSP) auction format in 2002, and most

other search engines also use GSP. Facebook adopted a different format, the Vickrey-Clarke-Groves (VCG) in 2010 (Bailey, 2015), as have some other platforms, including Google's recent AdSense. As will be explained later, the two formats differ in their rules that determine how much bidders pay for the slots that they win.

Which format is most efficient at auctioning ad space? Which captures more revenue for the platform owner? The present paper reports a human subject laboratory experiment designed to answer those questions.

One might think that such questions have already been answered, given the huge stakes and the vast amount of data held by companies hosting online ads. Unfortunately it is very difficult to compare the formats directly. For example, Google's AdWords users (as in Figure 1) are very different than their AdSense users, who place ads on their own websites. One could imagine conducting a field experiment with the format switched back and forth in some balanced fashion, but unannounced switches would probably provoke a lawsuit. Announced switches likely would cause backlash, since most major advertisers are comfortable with their strategies for a familiar format and would not welcome a short-term change in the rules.

Theory can provide insight, but it is not easy to model continuous time auctions for many different queries and heterogenous users. Two prize-winning papers — Varian (2007) and Edelman, Ostrovsky and Schwarz (2007) — make drastic simplifications and reach striking conclusions. As explained in Section 2 below, both papers model each auction as a static game of complete information with a fixed number of ranked slots and bidders. Simple payoff functions for bidders capture the difference between the two auction formats. The VCG version of the model has a unique Nash equilibrium; it is 100% efficient and splits the surplus in a particular way between bidders (the advertisers) and seller (the platform). The GSP version has multiple equilibria, all 100% efficient. Both papers highlight the GSP equilibrium with least seller revenue; although equilibrium bids are different, the outcome is the same as in the unique equilibrium of the VCG model. Thus the two formats may be revenue-equivalent.

Three papers report laboratory experiments informed by those theoretical models. Fukuda, et al. (2013) include detailed visually-oriented instructions describing the VCG pricing algorithm. Their subjects know their own absolute values for slots (referred to below as value per click, or VPC) but not those of other bidders, and every period they receive feedback on

own profit as well as all bids and payments. This first study finds no significant difference between the two formats, neither in efficiency nor revenue capture, but it considers only one parameter set and has relatively noisy data. Che, et al. (2013) consider only the GSP format. They use a minimal number of ad slots and bidders in discrete time, but test a variety of parameter sets and use a stringent empirical definition of efficiency. Results in their static complete information treatments closely parallel those in their dynamic incomplete information treatments. Noti, et al. (2014) compare VCG and GSP, using only one generic set of parameters and rapid discrete time (one auction per second). Even more than the previous studies, they find considerable overbidding in both auction formats, resulting in similar levels of inefficiency.

Section 3 lays out our laboratory procedures, which seek to incorporate the best aspects of the previous studies. We run parallel sessions with GSP and VCG formats, use moderate numbers of bidders (4) and slots (3), and a variety of more or less competitive parameter sets. Our user interface is even more visual and less text-oriented than that of Noti et al. Most important, the auctions are conducted in essentially continuous time with good feedback. As argued elsewhere, e.g., Pettit et al (2014), this enables human subjects to settle relatively quickly into long-run settled behavior. Of course, continuous time auctions are also an important realistic complication relative to the 2007 models.

Section 4 presents the results. As to the first question, we find that efficiency (under the stringent empirical definition) is far less than 100%, albeit higher than seen in previous studies with comparable parameters. In baseline treatments, efficiency is about 3% higher in VCG than in GSP, and in more competitive treatments is 10-20% higher. Most of these format differences are statistically significant. As to the second question, we find that the revenue captured by the platform is very similar across the two formats: a bit higher in VCG in some treatments and a bit lower in others, but the differences seldom are statistically significant. Average revenue capture is not far from that predicted by the static equilibrium model. The section includes a more detailed analysis of bidding behavior and some interpretative remarks.

Section 5 summarizes the findings, points out some practical implications, and notes a few caveats and directions for further research. An on-line Appendix includes background information on ad auctions, supplementary data analysis, and a copy of the instructions seen

by the subjects.

2 Theoretical Perspectives

Insight can be gained from the following simplified static framework, adapted from Edelman et al. (2007) and Varian (2007). The platform owner allocates slots $i = 1, ..., N$ to advertisers who bid at auction; e.g., $N = 7$ in Figure 1. All bidders know each slot's relative value, or click through rate (CTR), denoted α_i; these are sorted so that $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_N \geq 0$. Bidders $k = 1, ..., K$ may differ in their personal value per click (VPC), denoted s_k. Given bid profile $b = (b_k, b_{-k})$, an auction using format $F \in \{\text{GSP, VCG}\}$ allocates slot $i(k|b, F)$ to bidder k, who pays the platform owner an amount $r_k^F(b)$ specified below. Thus bidder k receives payoff

$$\pi_k = s_k \alpha_{i(k|b,F)} - r_k^F(b). \tag{1}$$

Both auction formats we consider allocate slots according to the bid ordering: the highest bidder gets slot 1, second highest gets slot 2, ..., and the Nth highest gets the last available slot N. That is, $i(k|b, F) = i(k|b) =$ the order rank (from highest to lowest) of bid b_k in $b = (b_1, ..., b_K)$. To avoid notational glitches, we assume that $K \geq N$ and set the slot CTR α_i (and the corresponding payment) equal to zero for $i > N$.

The surplus generated by an allocation of advertisers to slots is the overall value sum $S = \sum_{k=1}^{K} s_k \alpha_{i(k|b,F)}$, and an allocation is *efficient* when surplus is maximal. It is easy to see that efficiency results from allocating the best slots to advertisers with highest values per click, i.e., from an assortative match between α_i and s_k. The total payment $\sum_{k=1}^{K} r_k^F(b)$, also called the *revenue capture*, transfers some of the surplus to the platform owner.

Generalized Second Price (GSP). To streamline notation, renumber the bidders in decreasing order of bid, breaking ties randomly. Thus bidder k is the one with k^{th} highest bid, and so gets slot $i = k$. Under auction format $F = \text{GSP}$, her payment r_k is the least that allows her to retain position k, and in that sense generalizes the classic second price auction. More explicitly, bidder k pays b_{k+1} for every click she receives, so

$$r_k^{GSP} = \alpha_k b_{k+1}. \tag{2}$$

5

Auctions are traditionally analyzed as games of incomplete information in which bidders know only the distribution of rival bidders' values. However, for reasons explained by Varian (2007) and Edelman et al. (2007), some of which are alluded to below, it is useful here to focus on the game of complete information in which each bidder k knows rivals' values $s_j, j \neq k$ as well as her own value.

Those authors find a range of symmetric Nash equilibrium (SNE) bidding functions $b_k = B(s_k, s_{-k}, N, K)$ for the game defined by payoff function (1, 2). That range is characterized by a system of inequalities that state that each bidder prefers neither to outbid the next higher bidder nor to underbid the next lower bidder. All SNE are efficient, because the bid functions are increasing in the relevant sense, and therefore induce assortative allocations and maximize surplus. But different SNE divide up that surplus differently between advertisers and the platform.
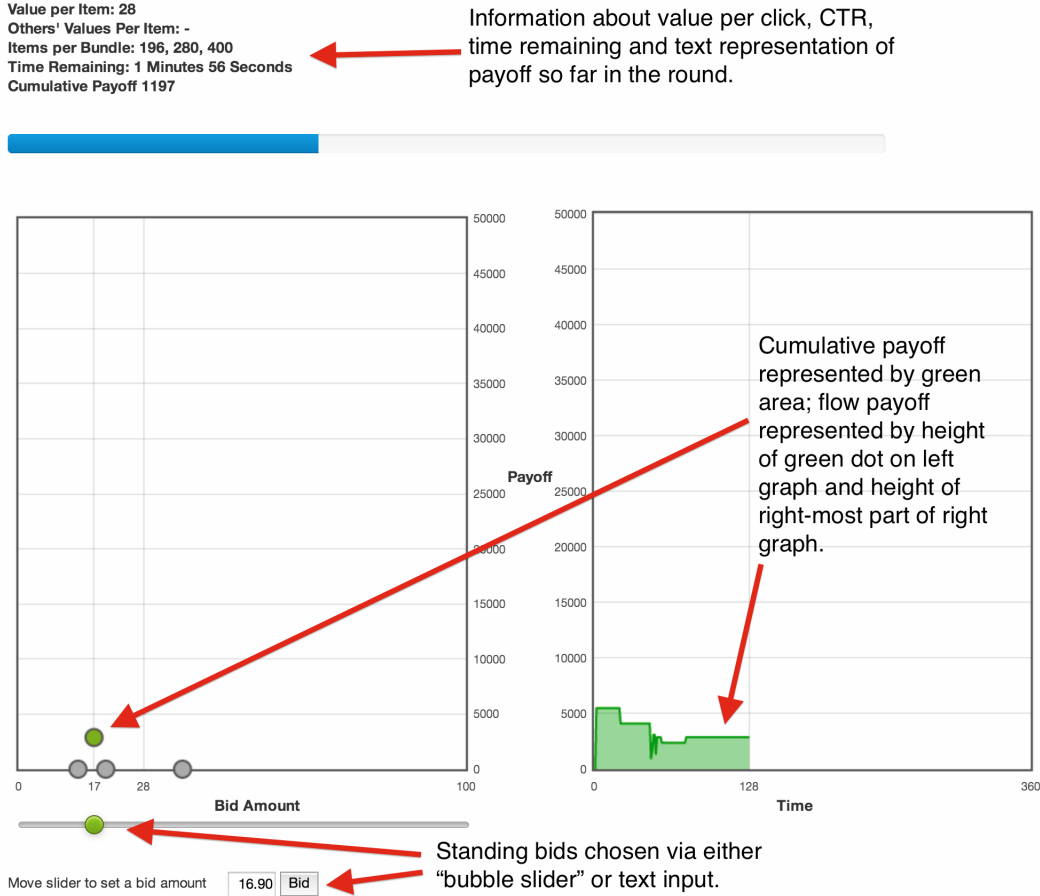
Vickrey-Clarke-Groves (VCG). Under the VCG format each bidder pays the revealed value displaced by his participation in the auction. To spell it out using the streamlined notation (bidder indexes sorted from highest bid to lowest), bidder k pays

$$r_k^{VCG} = \sum_{j=k}^{N} (\alpha_j - \alpha_{j+1}) b_{j+1}. \tag{3}$$

The idea is that only lower bidders $k+1, ..., N$ are displaced by k, and each is bumped down one slot and so loses the difference $(\alpha_j - \alpha_{j+1})$ in slot CTRs times b_{j+1}, the VPC revealed in his bid. Varian (2007) and Edelman et al. (2007) note that, as a special case of Leonard (1983), there is a unique NE of of the game defined by the payoff function (1, 3). Truthtelling $(b_k = s_k)$ is weakly dominant and, of course, efficient.

The 2007 articles both also show that the allocation and payments in this equilibrium of the VCG auction coincides with an extreme SNE of the GSP auction, the one with lowest payments. The articles suggest that that extreme SNE is the most attractive prediction of the GSP auction, implying that the two auctions, despite inducing different bid functions, will both be fully efficient and will capture the same revenue for the platform.

Figure 2: User Interface.

Value per Item: 28
Others' Values Per Item: -
Items per Bundle: 196, 280, 400
Time Remaining: 1 Minutes 56 Seconds
Cumulative Payoff 1197

Information about value per click, CTR, time remaining and text representation of payoff so far in the round.

Cumulative payoff represented by green area; flow payoff represented by height of green dot on left graph and height of right-most part of right graph.

Bid Amount

Move slider to set a bid amount   16.90 Bid

Standing bids chosen via either "bubble slider" or text input.

3   Experiment Design

Our experiment implements a dynamic version of the simple static model just described. Each ad auction consists of $K = 4$ human bidders competing for $N = 3$ slots, and their take-home payments are governed by equations (1, 2) in GSP trials, and by equations (1, 3) in VCG trials. In the static model, bids are chosen once and for all, and are submitted simultaneously. By contrast, in our experiment as in the field, bidders can adjust their bids freely in real time. The payment equations determine the instantaneous flow payoff, and players accumulate take-home earnings continuously throughout each market period.

Figure 2 is a snapshot of the computer screen faced by a human player, with comments overlaid. Text in the upper left of the screen includes payoff-relevant information such as the player's VPC, here $s_k = 28$, and the CTRs of the three slots (here $(\alpha_1, \alpha_2, \alpha_3) =$

7

(400, 280, 196). Following standard practice of using neutral language (to avoid triggering subjects' preconceptions about ads), the VPC is referred to as referred to as "value per item," and the CTR as "items per bundle." Note that players are not told their rivals' VPCs.

The box on the left of the screen shows the current bids of all four players on a horizontal scale; the three gray dots show the rivals' bids, currently approximately 14, 20 and 40. The player can adjust her own bid whenever she wants by dragging the slider (or typing in a number) below the box; her green dot follows. The height of that dot represents her flow payoff rate, with the scale displayed on the vertical axis. The green area in the box on the right shows the payoff accumulated so far. The snapshot is at tick 128 of 360, i.e., 64 seconds into a 180-second period. (As explained in Pettit et al. (2014), the software has actual latency on the order of 50 ms, but here we set the data capture ("tick") rate at 500 ms.) Negative payoffs are represented by red area in right box and a bold red arrow near the left box alerts subjects that their flow payoff is below the x-axis. A screen shot of this case, and complete instructions, can be found in Online Appendix C.

Phase I of our experiment compares GSP sessions to paired VCG sessions with the same parameter set. We begin with baseline parameters, similar to those used in previous studies, which give each slot 70% of the CTR as the next better slot, and spread the bidders' VPCs quite widely and evenly. We use two versions of the VPC schedule to check robustness and to avoid boredom and excessive familiarity with other subjects' possible values. We then consider two versions of more competitive parameters (Competitive1.1 and Competitive1.2) in which bidders' VPCs are more tightly bunched and the CTRs of the top two slots are not far apart. We also introduce a second set of competitive parameters (Competitive2.1 and Competitive2.2) in which the second and third slots are worth far less than the top slot. The schedules are shown in Table 1.

Phase II checks robustness to instructions. Understanding the payment rules, especially for VCG, is a real challenge for non-economists, but previous investigators developed excellent pedagogical tools that we were able to refine. Nevertheless, it is reasonable to wonder whether the differences observed across formats in Phase I might be caused by differences in how well subjects understood instructions rather than by strategic differences in the formats. Therefore, in Phase II we used identical streamlined instructions for both formats — subjects were simply told that the rank of their *own* bid determines which bundle (slot) they

8

Parameter Set	VPC	CTR	Phase: Sessions Run
Baseline 1.1	$s_k \in \{40, 30, 20, 10\}$	$\alpha_i \in \{400, 280, 196\}$	Phase I: 2 VCG, 2 GSP
Baseline 1.2	$s_k \in \{37, 28, 22, 13\}$		
Competitive 1.1	$s_k \in \{47, 45, 40, 37\}$	$\alpha_i \in \{400, 380, 190\}$	Phase I: 2 VCG, 2 GSP
Competitive 1.2	$s_k \in \{39, 37, 34, 30\}$		Phase II: 2 VCG, 2 GSP
Competitive 2.1	$s_k \in \{47, 45, 40, 37\}$	$\alpha_i \in \{400, 160, 120\}$	Phase I: 1 VCG, 1 GSP
Competitive 2.2	$s_k \in \{39, 37, 34, 30\}$		Phase II: 2 VCG, 2 GSP
			Phase III: 2 VCG, 2 GSP

Table 1: Experiment Design. All sessions lasted 20 periods with 8 subjects whose assignments reshuffled randomly each period. Each period lasted 180 seconds and featured two markets, each with K = 4 subjects competing for N = 3 ad slots. Parameter set x.1 shown in the second column was used in 10 periods and parameter set x.2 in the other 10 periods. The last column shows how many sessions of each phase and format used each parameter set pair.

win, and that *other players'* bids determine the cost of that bundle.

Phase III is a different sort of robustness check. As we will soon see, with either format subjects were consistently able to converge to a "behavioral equilibrium" within a few periods. But, having settled down, can subjects then adjust to a new format and, if so, how long does this process take? In Phase III we used the identical streamlined instructions, ran the first 10 periods in one format, warned the subjects that the (still unspecified) rules determining the cost of the bundle were about to change once and for all, and then ran the remaining 10 periods in the other format.

Except for the first few Phase I sessions (which did not use Competitive2 parameters), each session was assigned a treatment randomly. Upon arrival, subjects read through the text instructions and then the conductor read them aloud, with supplementary slides on the payoff and cost rules. After Q & A, a short quiz and a five minute unpaid practice period, subjects played 20 paid three minute periods. Each session consisted of 8 subjects, playing in two groups of four. Group membership, VPC assignments within the group, and the parameter sets (x.1 or x.2) were reshuffled each period; see Table 1 for specifics. The typical session lasted less than two hours, and subjects earned $23.84 USD on average.
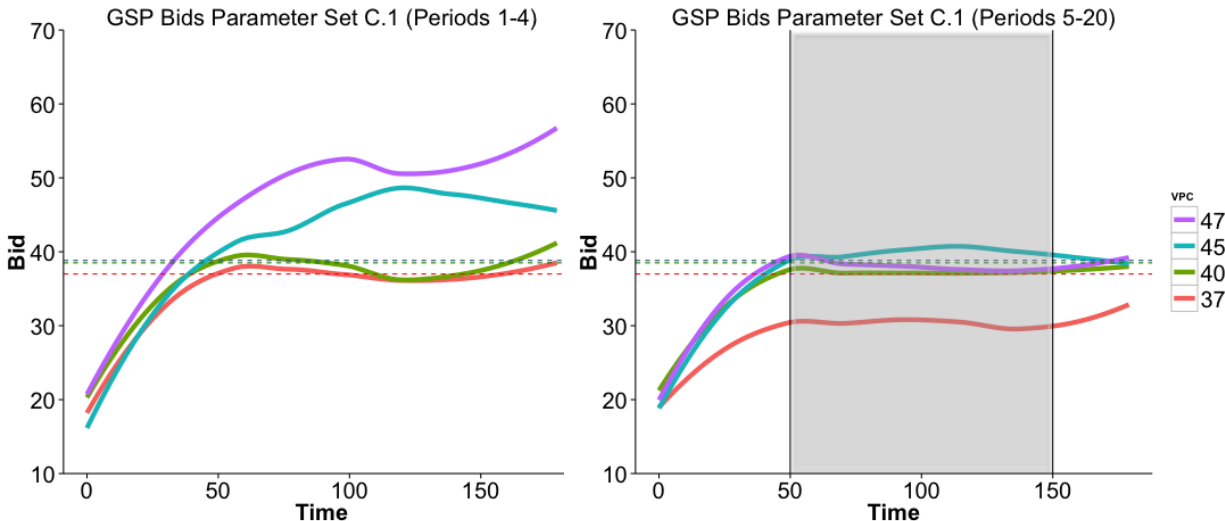
4 Results

Figure 3: Bids and Sample Selection. Mean bids by VPC assignment for indicated periods and treatments, smoothed over time via a locally weighted polynomial regression (LOESS with $\alpha = 0.75$). Subsequent analysis focuses on the data in the shaded area between the two vertical lines.

Figure 3 summarizes bidding behavior in Phase I GSP sessions for one of our parameter sets. The bids of each player are sampled every half-second, averaged for each time sample across all players in the same VPC role (here either 47, 45, 40 or 37, from parameter set C.1) for all indicated periods in all those sessions, and smoothed over time. The first panel shows that, relative to GSP equilibrium, overbidding is rampant in periods 1-4, especially for the two players with highest VPCs. The second panel shows that bidding is much more moderate in periods 5-20. Of course, an upward trend remains in the first minute or so of a typical period since bids are initialized to zero, and there may also be some end-of-period effects.

Both panels of Figure 3 are typical in that (a) behavior in the first few periods of most sessions includes overbidding reminiscent of that seen throughout most previous lab studies of ad auctions, and (b) bids in subsequent periods show little trend after about 40-60 seconds have elapsed until perhaps the last 20 seconds. Our interpretation is that, with our continuous time user interface, a few periods of experience enables humans to understand

10

the strategic environment well enough to make thoughtful and consistent responses. They also need some time at the beginning of each period to settle down into their new roles and for the influence of the initialization to wear off. For those reasons, the analysis to follow will cover data only from periods 5-20 and $t \in (50, 150) \subset [0, 180]$ seconds.

	Efficiency	Revenue
Constant	0.827***	1.029***
	(0.027)	(0.077)
VCG	0.089***	−0.061
	(0.029)	(0.068)
Competitive1	−0.236***	−0.084
	(0.057)	(0.087)
Competitive2	−0.135***	−0.120
	(0.038)	(0.098)
Phase II	−0.138***	−0.090
	(0.040)	(0.085)
Phase III - GSP to VCG	−0.037	−0.048
	(0.034)	(0.080)
Phase III - VCG to GSP	−0.031	−0.065
	(0.021)	(0.080)

Table 2: Results Overview. Entries are estimated coefficients (and robust standard errors) for Efficiency and Revenue Capture regressed on treatment dummy variables. Reference (omitted) dummies are Baseline and GSP. Significance at levels $p < 0.10, 0.05, 0.01$ are indicated by *,**,***.

4.1 Efficiency

Efficiency is calculated as $\text{Eff} = \frac{AS-RS}{MS-RS}$ where AS, RS, and MS respectively are actual surplus, random surplus, and maximum possible surplus. Recall that surplus $S = \sum_{k=1}^{K} s_k \alpha_{i(k)}$ is maximized by an assortative allocation of players with highest VPC s_k to slots with highest CTR α_i; here RS is the mean surplus over all possible allocations. The actual surplus is computed from the 200 sampled allocations between $t = 50$ seconds and $t = 150$ seconds in each period from 5 to 20 each session. Note that $\text{Eff} < 0$ for a disassortative allocation.

The Efficiency column of Table 2 displays regression results using data from all treatments. Overall, GSP achieved 83% efficiency on average, while VCG achieved about 9% greater efficiency. Relative to baseline parameters, sessions run with Competitive1 and Competitive2 parameters on average saw 24% and 14% lower efficiency rates, respectively. Phase II treatments, using the streamlined instructions, also reduced efficiency about 14% relative to baseline, and Phase III treatments reduced efficiency 3-4 %.

	Comp1	Comp2	Competitive	Baseline	Pooled
Phase I					
Constant	0.535***	0.646***	0.555***	0.862***	0.748***
	(0.006)	(0.000)	(0.024)	(0.002)	(0.080)
VCG	0.212***	0.183***	0.210***	0.027***	0.114
	(0.020)	(0.000)	(0.030)	(0.004)	(0.084)
Phase II					
Constant	0.665***	0.646***			0.655***
	(0.020)	(0.151)			(0.070)
VCG	0.116***	0.196			0.156**
	(0.029)	(0.152)			(0.073)
Phase III					
	GSP to VCG	VCG to GSP			
Constant	0.841***	0.788***			0.815***
	(0.010)	(0.016)			(0.017)
VCG	−0.013	0.106***			0.046
	(0.099)	(0.004)			(0.053)

Table 3: Efficiency coefficient estimates (and robust standard errors). Reference (omitted) dummy is GSP. Significance at levels $p < 0.10, 0.05, 0.01$ is indicated respectively by *,**,***.

More precise format comparisons appear in Table 3 and Figure 4. In the Phase I baseline sessions, the Table shows mean efficiency of $86.2 \pm 0.2\%$ under GSP and about 3% higher under VCG, a statistically significant but economically small difference. The difference is economically much larger with competitive parameters: VCG achieves 21% higher efficiency than GSP, with both Competitive1 and Competitive2 parameter sets contributing. The first panel of the Figure shows that, pooling across all parameter sets, the difference in Phase I efficiency comes mainly from a much smaller fraction of very inefficient allocations (with Eff < 0.5) under VCG and a correspondingly greater fraction of maximally efficient allocations.
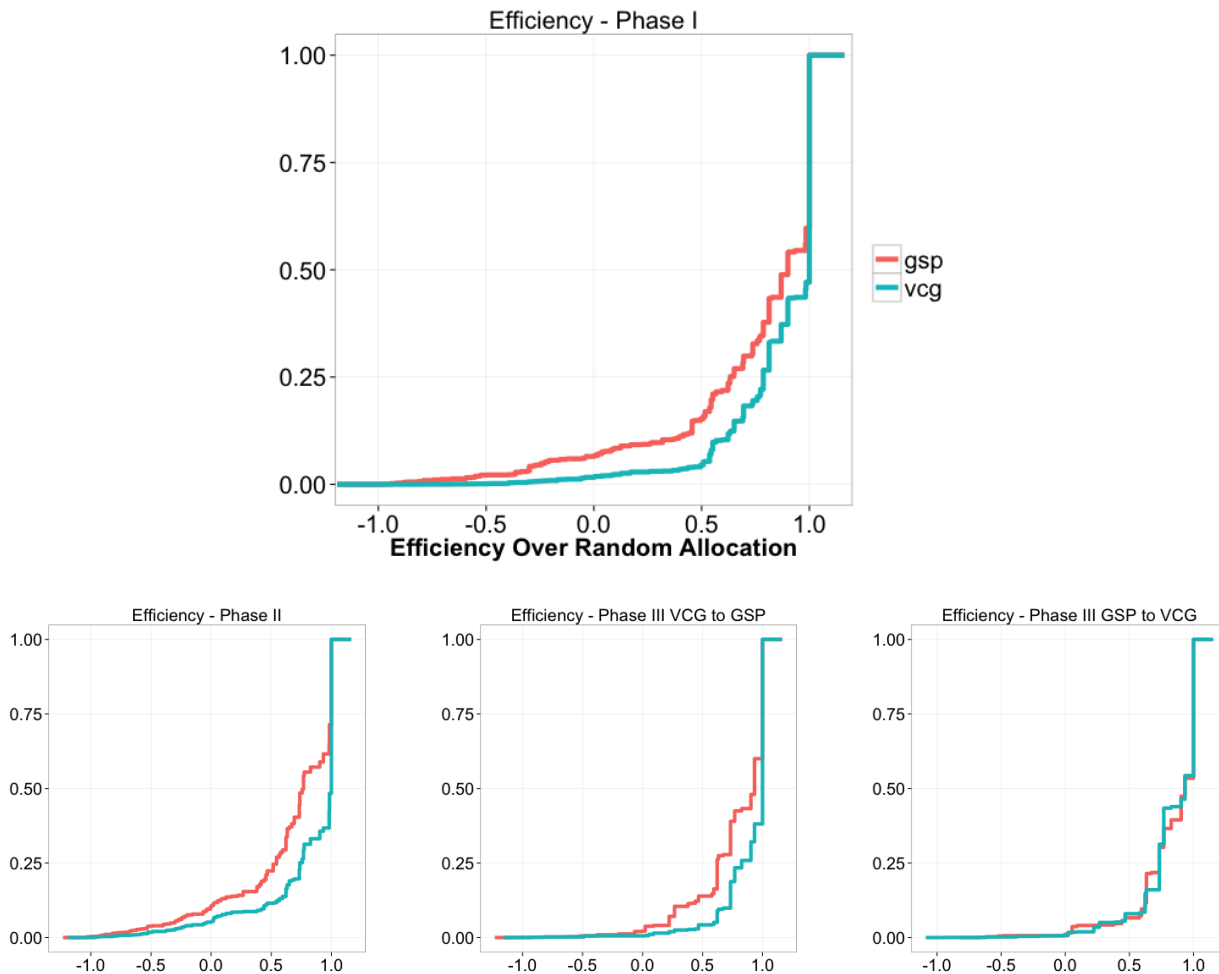
Figure 4: Empirical Cumulative Distribution Functions for Efficiency.

Phase II treatments tell a similar story: a 16% higher mean efficiency under VCG overall, 12% higher under Competitive1 and (statistically insignificantly) 20% higher under Competitive2. Phase III sessions that start with GSP show no difference in the efficiency mean or distribution across formats, but the Phase III sessions that start with VCG show a significant drop in efficiency, about 10% lower (paired t-test p-value = 0.040 and paired Wilcoxon test p-value = 0.038) after the switch to GSP.

13

4.2   Revenue Capture

We measure revenue captured by the platform as a fraction of the level predicted in the unique VCG equilibrium. The second column of Table 2 shows no significant treatment effects for this measure. Relative to the GSP baseline, the VCG format on average captures 6% less revenue and the Competitive2 parameters lead to 12% less capture, but the standard errors are about the same magnitude.

	Comp1	Comp2	Competitive	Baseline	Pooled
Phase I					
Constant	0.904***	0.845***	0.894***	1.087***	1.015***
	(0.014)	(0.000)	(0.014)	(0.078)	(0.076)
VCG	0.008	0.112***	0.023	−0.134	−0.075
	(0.014)	(0.000)	(0.015)	(0.149)	(0.108)
Phase II					
Constant	0.938***	0.869***			0.904***
	(0.002)	(0.117)			(0.056)
VCG	−0.024	0.046			0.011
	(0.053)	(0.118)			(0.062)
Phase III					
	GSP to VCG	VCG to GSP			
Constant	0.965***	0.936***			0.951***
	(0.005)	(0.020)			(0.012)
VCG	−0.030**	−0.005			−0.017*
	(0.015)	(0.010)			(0.010)

Table 4: Revenue Capture as % of VCG Equilibrium.Entries show coefficient estimates (and robust standard errors). Reference (omitted) dummy is GSP. Significance at levels $p < 0.10, 0.05, 0.01$ indicated respectively by *,**,***.

The more precise estimates in Table 4 and the empirical distributions shown in Figure 5 reinforce the message that the formats are nearly revenue equivalent. Both formats capture close to 100% of predicted revenue in Phase I sessions, about 90% in Phase II and about 95% in Phase III. The most significant treatment coefficient in the Table suggests that VCG enables 11.2% more revenue capture than GSP with Competitive2 parameters in the Phase I sessions, but this is not confirmed in Phase II or in pooled data. The other significant

coefficient (at $p < 0.05$) suggests that VCG enables 3% less revenue capture when subjects encounter this format following a ten period experience with GSP.
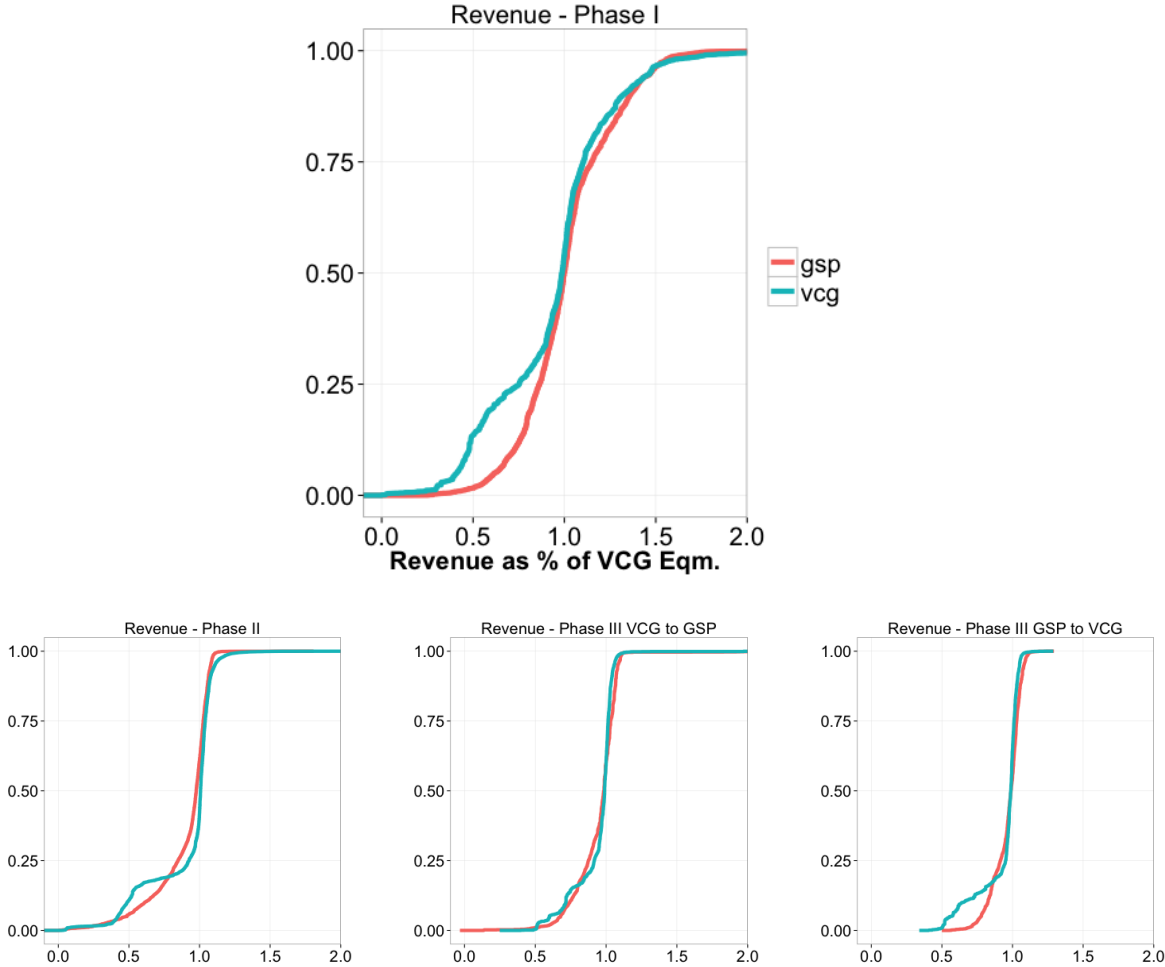

Figure 5: Empirical Cumulative Distribution Functions for Revenue Capture.

For the most part, Figure 5 shows empirical distributions for the two formats that follow each other quite closely. A striking exception is that, in Phases I and II and in half of Phase III, there is a hump in the VCG distribution. It indicates that the lowest quartile of revenue capture is worse in VCG than in GSP.

What accounts for that shift? Can we trace it to bidding behavior? The next subsection investigates.
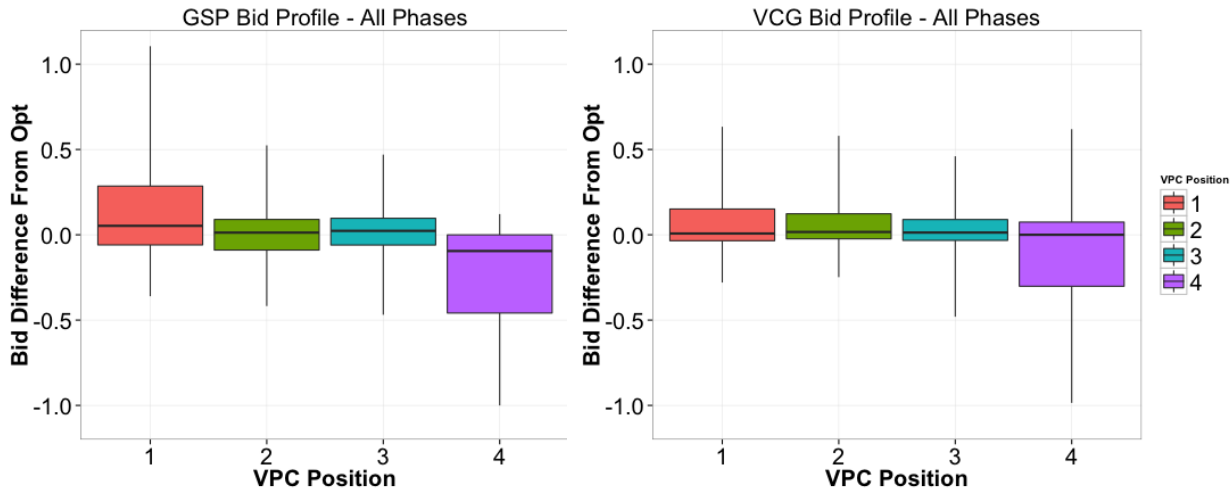
Figure 6: Bid Deviations by bidder position (or role). The color-coded boxes show the interquartile range, the black line the median, and the vertical line segments span the 5% and 95% percentiles.

4.3 Bidding Behavior

Figure 6 shows the distribution of deviations of actual bids from those predicted in the unique VCG equilibrium or in the outcome-equivalent GSP equilibrium. The data are disaggregated by position in the value-per-click schedule (1 being the highest VPC assigned and 4 being the lowest), normalized by the predicted bid, and aggregated across all subjects, all phases and all (non-format) treatments.

Theory constrains bids from Position 1 only to be above the second highest bid; the upward deviations are inconsequential. Bidders in positions 2 and 3 have the greatest impact on efficiency, since the slot allocation is sensitive to both upward and downward deviations from the equilibrium prediction. Perhaps for that reason, observed deviations are smallest for these positions; medians tend to be at or very slightly above prediction, and deviations go both ways.

In equilibrium in both formats, players in position 4 bid their true VPC and earn nothing. Deviating to a higher bid might cause an out-of-pocket loss, while deviating to a lower bid sacrifices very little expected profit (since there is none in equilibrium) and never risks an out-of-pocket loss. Therefore it is not surprising to see that more than 75% of observed

bids from this position in GSP are below prediction. In VCG, about 50% of such bids are below prediction, and the distribution is skewed: most positive deviations are quite small, while negative deviations often substantial and bidding zero (for a deviation of -1.0) is not uncommon.

4.4 Discussion

The consequences of such low bidding differ for the two formats. The VCG payment rule is recursive, and low bids from position 4 have a ripple effect on revenue paid by all bidders. By contrast, in GSP such bids have a direct impact only on the next higher bidder.

We see this difference as a likely explanation for the hump seen in the cumulative distribution functions for VCG revenue capture. The lowest bids from position 4 take a much bigger bite out of revenue capture in VCG than they do in GSP, and thereby give the VCG distribution a fatter lower tail.

Overall, revenue capture in VCG is very close to that in GSP, so it seems reasonable to look for an offsetting effect. The obvious candidate is the greater efficiency of VCG. Other things equal, a bigger total surplus gives more to everyone, including the platform.

But why is VCG more efficient than GSP? Section 4.1 documented a roughly 10% improvement in exact assortative allocations for VCG and a corresponding reduction in very inefficient allocations relative to GSP. Could that be an artifact, perhaps just that VCG instructions somehow are presented better? That explanation seems unlikely since the regularity persists in Phase II, in which instructions are the same for both formats.

We believe that VCG's superior efficiency is intrinsic to the auction rules. Efficiency is always a matter of allocation. In both formats, low bidding might impair revenue capture, but it is consistent with the 100% efficient assortative allocation as long as nobody underbids a rival who has a lower VPC. Of course, such underbidding is impossible from position 4, but it is quite possible from the other positions.

The VCG format, in practice as well as in theory, seems to encourage players to bid fairly close to their true VPCs. By contrast, in the GSP predicted equilibrium, players in positions 1-3 are supposed to shade their bids below true VPC, and it is easy to imagine

some players shading bids so much as to underbid a rival with lower VPC.

Figures in the Appendix suggest a refinement of this insight. VCG analogues of Figure 3 remind the reader that the "truth-telling" predicted bids in VCG are as widely separated as the VPC schedule permits. By contrast, as in Figure 3, in GSP the predicted bids tend to be tightly bunched. Thus GSP bidding has a smaller margin of error before efficiency suffers.

Our analysis focused on behavior and outcomes in the time interval [50, 150] seconds of periods 5-20 of all sessions. The Appendix shows that other reasonable choices of time intervals and periods lead to very similar conclusions.
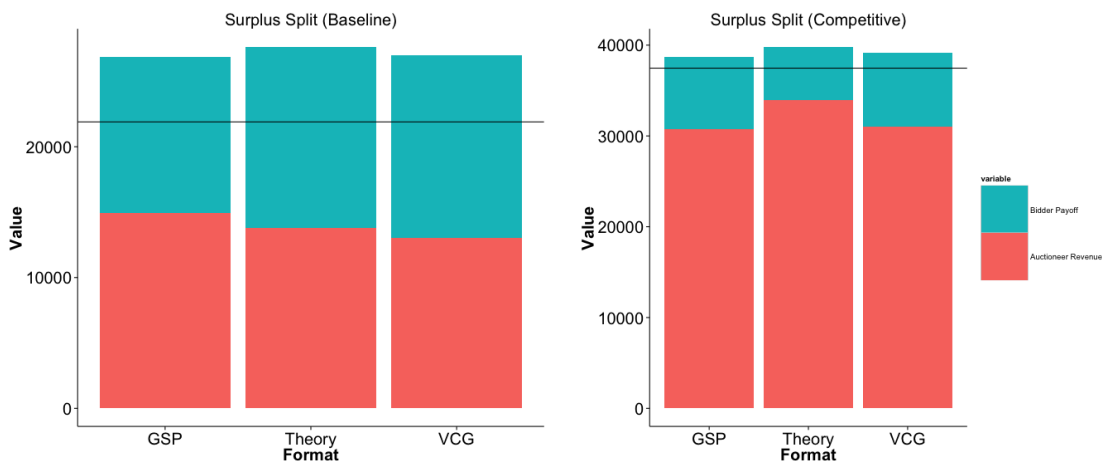

Figure 7: Surplus Split. The height of each bar reflects the total surplus, the horizontal line marks expected surplus in a random allocation, and the bar colors show the split predicted by theory and that realized on average in Phase I data in the two formats.

Figure 7 provides a helpful perspective on much of our data. Panel A shows that with the baseline parameter set, the GSP sessions and the VCG sessions both fall slightly short of the maximum possible total surplus; of course, the shortfalls loom larger when they are measured as surplus above that in a random allocation (black line). The differences in revenue capture (red portion of the bar) seem large, but we see from Table 3 that that GSP revenue averaging 8.7% above prediction and VCG revenue averaging $13.4 - 8.7 = 4.7\%$ below, differ insignificantly from the prediction, due to considerable sampling variability.

Panel B of the figure summarizes comparable data from the more competitive parameter sets. Again actual total surplus falls only slightly short of the theoretical maximum, but the

shortfalls in surplus above the random allocation are more striking — a 44% shortfall with GSP versus a 23% shortfall with VCG, as can be verified from Table 4. Revenue capture is quite similar for the two formats, about 11% below prediction in GSP versus about 9% below in VCG; again see Table 3.

5   Conclusion

So which format is better, Vickrey-Clarke-Groves (VCG) or Generalized Second Price (GSP)? Simplicity is important, but that razor cuts both ways. It is definitely easier to explain the GSP payment rule (pay the next lower bid per click) but, on the other hand, VCG equilibrium bids (bid true value per click) are as simple as can be. As we saw in Section 2, theory suggests that the formats are equally good. In the unique equilibrium of the VCG model and in a focal equilibrium of the GSP model, the platform captures the same revenue, and all equilibria are 100% efficient.

Our laboratory results largely support the theory but add crucial nuance. We fail to reject the null hypothesis of revenue equivalence across formats, but we do reject the hypothesis of equal and maximal efficiency. Roughly speaking, we find that VCG produces a bigger pie, but GSP awards a larger fraction of its pie to the platform, so that it ends up with about the same amount of revenue under either format.

How robust are these results? Relative to the theory in Section 2 (and relative to preceding laboratory investigations), we we able to incorporate several realistic complications that could make a difference — real time action, a variety of parameter configurations, and a highly visual user interface focused on profitability. We believe that our subjects adapted themselves well to the strategic environment, and that the deviations from equilibrium behavior that we detected are likely to persist in settings yet more realistic than ours.

What are the practical implications? If indeed the two formats capture the same revenue, why should the platform owner prefer one over the other? The preceding Discussion suggested that the empirical revenue equivalence arose from offsetting effects, so it seems possible that some environments might upset the balance and favor one format over the other. We believe that VCG is generally the better choice for a new platform, because in

the long run efficiency matters for revenue capture as well as for its own sake. Customers (advertisers) do better under VCG, so they are more likely to be attracted to and stick with a platform that uses this more efficient format. On the other hand, it is not clear whether anyone benefits if an established GSP platform switches to VCG. As suggested in our Phase III sessions, the transition costs could be substantial — advertisers would have to learn to increase their bids, and while they did, revenue capture and efficiency might both suffer (Varian and Harris, 2014).

We close with a caveat and call for further research. Appendix A notes several complications omitted from our analysis, e.g., non-separable slot valuations, advertising budget constraints, and reserve prices for slots. We now see no reason why any of these complications would reverse our findings, but that question can only be answered by new empirical research. We hope that our work so far inspires further laboratory experiments, and possibly small field experiments. Conversely, we hope that the regularities seen in our data inspire further theoretical investigations into bidding behavior under the two ad auction formats, and perhaps even the invention of new formats.

Appendix A   Institutional Background

Since the late 1990s, the majority of online ad space has been allocated through an auction mechanism whereby advertisers pay a cost every time a user "clicks" on the displayed ad[1]. This practice was first instituted in 1997 when Overture (then GoTo) introduced a new model for selling online advertising in which advertisers could target their audience more accurately. The mechanism used in early ad auctions was a generalized first price auction. While this method was more efficient and provided advertisers with better targeting options than the cost-per-impression negotiated rates that had been used prior, the first price mechanism was still inefficient and advertisers exhibited bidding behavior that led to substantially inefficient allocations. Furthermore, this mechanism was shown to reduce platform revenues by as much as a 10% when compared to a Vickrey auction (Ostrovsky and Schwarz, 2009). This led to the formulation of a new mechanism, the Generalized Second Price auction[2].

GSP has since become the dominant mechanism used for allocating ad space in this realm. This is the format that is currently employed by Google[3], Bing, and Yahoo![4]. Facebook, however, chose to use the Vickrey-Clarke-Groves mechanism to allocate ad space. This seems a natural choice for Facebook, as many of the ads displayed on the Facebook platform do not follow the more common "stacked" format and VCG is more generalizable to different ad position configurations.

The main theory papers describing these two mechanisms simplify many of the unique complications that advertisers face in this market. Features such as throttling/pacing, quality adjustments, and possible combinatorial environments are just a few complications which may affect bidders' strategy as well as format performance. While these complications are outside of the scope of our experiment, they are useful to acknowledge in order to better understand the mapping between the results of this paper and behavior in online ad auction markets.

---

[1]While the cost-per-click payment model is used most often, cost-per-impression and cost-per-acquisition are also common

[2]GSP was first instituted by Google in 2002

[3]Google uses GSP as the mechanism for allocating ads for the majority of the ad space it sells but has been using VCG for a small set of ads called "contextual ads" since 2012. Varian and Harris (2014)

[4]Microsoft AdCenter runs the auctions that allocate ads for both Bing and Yahoo!

Throttling and pacing are features used by both Google and Facebook which adjust an advertiser's effective bid so that a daily budget is met at the optimal time (the end of the day). Facebook's website uses the analogy of a runner in a race: "sprint too early and risk fading away before the finish line, but sprint too late and never make up the distance. Pacing ensures uniform competition throughout the day across all advertisers and automatically allocates budgets to different ads."[5]. Quality multipliers or quality scores are a feature discussed in Varian (2007) whereby an advertiser's bid is adjusted based on that ad's performance relative to a baseline. These multipliers can increase or decrease an advertiser's effective bid depending on whether the ad performs better or worse, respectively, than the baseline ad on average[6]. Lastly, companies which are controlling many brands, with overlapping target audiences, may face a combinatorial auction in which there is a preference for ad positioning which is dependent on the ads it is shown alongside.

A recent paper by Goldman and Rao (2014) finds a minimal homogenous effect of ad slot position on click quality. The main results of that paper were that the separability of click through rates between ad slots is not as clean as has been suggested by much of the theory on these auctions. The effect that this would have on our experiment is minimal, given that this is an issue with the performance of ad slots rather than the allocation formats and should affect both VCG and GSP equally. The reader should take note, however, that this may have effects on the efficiency rates, which depend on a clear definition of ad slot expected CTR. The other papers mentioned do not have much effect on our experiment, so we leave it to the reader to explore these papers further.

Efficiency has value for the seller in more than just the direct costs described in the paper above. From a theoretical standpoint, it is well known that the expected revenues of the auctioneer are maximized when the following hold: (i) the expected bid of a bidder with value $s_k = 0$ is 0, (ii) only bidders with a positive "virtual valuation"[7] are allocated clicks, and (iii) among them, bidders with higher virtual valuations are allocated as many clicks as possible (Ostrovsky and Schwarz, 2009). One other possible ramification of a low efficiency rate is that disassortative matching may result in end users being much less likely to click on ads in the future. When a user clicks on an ad which gives a bad experience,

---

[5]https://developers.facebook.com/docs/marketing-api/pacing

[6]In the multiplier increases the advertiser's bid, they will pay at most the submitted bid.

[7]Virtual valuation $\psi_k = s_k - \frac{1 - F_k(s_k)}{f_k(s_k)}$

the probability that that user clicks on any ad in the future is drastically lower. The less efficient the mechanism is, the more likely that the end user is shown an irrelevant ad.

One striking result in our experiment is that in many of our treatments, subjects do not exhibit the systematic overbidding that is persistent in other lab experiments. This could be due to a number of factors, many of which are related to our particular experiment interface and the continuous time feature. For one, our interface may reduce the "joy of winning" and what is often called "auction fever" by displaying the flow profit rate and cumulative profit rate in an easy to interpret graphic and not explicitly showing who won which bundle[8]. Subjects may also be learning at a faster rate than in previous experiments due to the sheer number of auctions in which they are participating. There is evidence of this learning effect, as early rounds show systematic overbidding.

---

[8]Design features that display which ad slot a subject won may have contributed to some of the overbidding seen in Nisan (2014)

Appendix B   Bid Profiles
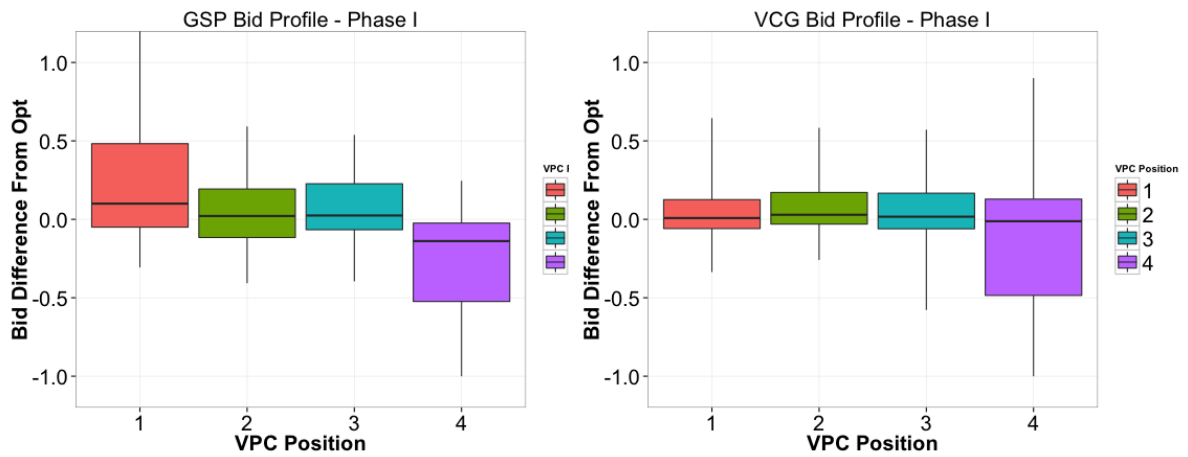
Figure 8: Bid Profiles Phase I
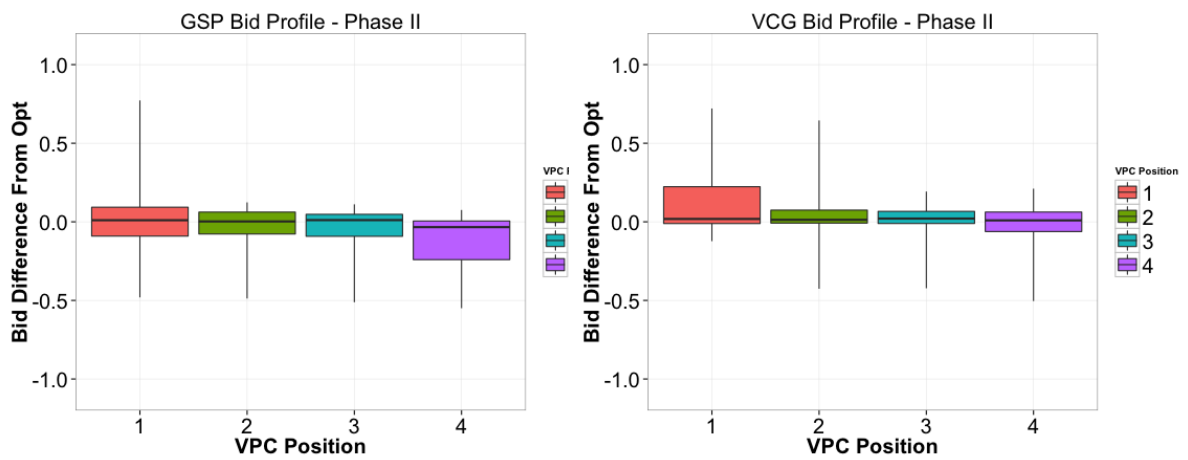

Figure 9: Bid Profiles Phase II

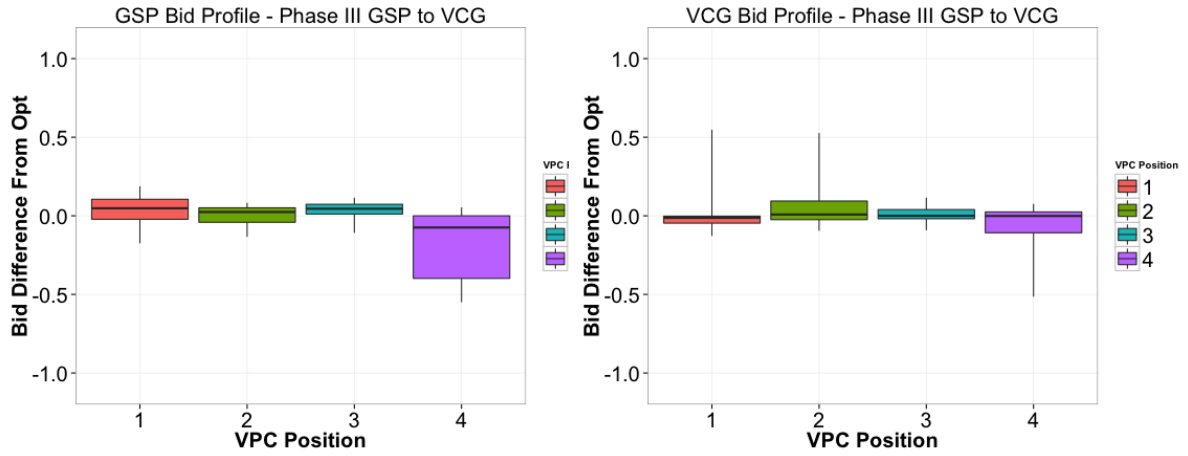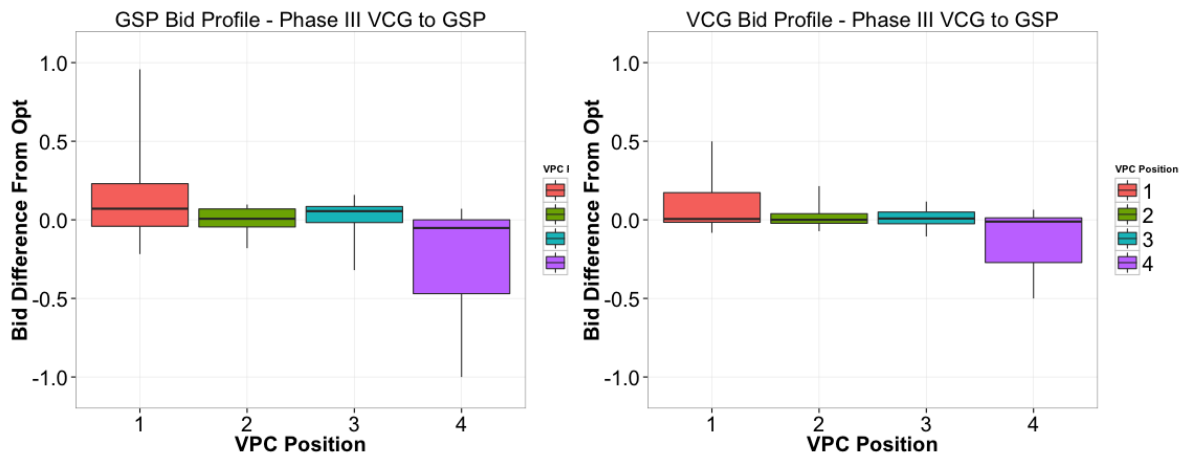Figure 10: Bid Profiles Phase III - GSP to VCG

Figure 11: Bid Profiles Phase III - VCG to GSP

Appendix C   Supplemental Regressions

Table 5: Baseline & GSP reference, robust standard errors

	Efficiency		Revenue	
	(1)	(2)	(3)	(4)
Constant	0.827***	0.827***	1.029***	1.029***
	(0.027)	(0.027)	(0.077)	(0.077)
VCG	0.089***	0.089***	−0.061	−0.061
	(0.029)	(0.029)	(0.068)	(0.068)
Competitive1	−0.236***	−0.215**	−0.084	−0.124
	(0.057)	(0.091)	(0.087)	(0.097)
Competitive2	−0.135***	−0.135***	−0.120	−0.120
	(0.038)	(0.038)	(0.098)	(0.098)
Same Instructions	−0.138***	−0.127*	−0.090	−0.100
	(0.040)	(0.073)	(0.085)	(0.089)
GSP to VCG	−0.037	−0.037	−0.048	−0.048
	(0.034)	(0.034)	(0.080)	(0.080)
VCG to GSP	−0.031	−0.031	−0.065	−0.065
	(0.021)	(0.021)	(0.080)	(0.080)
Parameter Fixed Effects	No	Yes	No	Yes

Note: *p<0.1; **p<0.05; ***p<0.01

Table 6: Efficiency; reference vars are *GSP* and *Baseline*. Clustered at the session level

	Efficiency:		
	(1)	(2)	(3)
vcg	0.110**		0.023***
	(0.055)		(0.005)
res_price		−0.099***	−0.105***
		(0.025)	(0.031)
same_ins		−0.150***	−0.214***
		(0.044)	(0.067)
g2v		−0.084**	−0.065***
		(0.035)	(0.007)
v2g		−0.074***	−0.109***
		(0.007)	(0.013)
comp1		−0.254***	−0.341***
		(0.080)	(0.013)
comp2		−0.147**	−0.224***
		(0.066)	(0.005)
vcg:res_price			0.019
			(0.046)
vcg:same_ins			0.133*
			(0.069)
vcg:g2v			−0.032
			(0.058)
vcg:v2g			0.077***
			(0.017)
vcg:comp1			0.213***
			(0.022)
vcg:comp2			0.159***
			(0.005)
Constant	0.740***	0.884***	0.869***
	(0.051)	(0.006)	(0.005)

Note:	*p<0.1; **p<0.05; ***p<0.01

27

Table 7: Efficiency – reference vars *GSP*, *baseline*, *base param1*, *comp2 param2*. Clustered at the session level

	Efficiency:				
	(1)	(2)	(3)	(4)	(5)
vcg	0.110**		0.023***	0.037***	0.023***
	(0.055)		(0.005)	(0.006)	(0.005)
res_price		−0.099***	−0.105***	−0.083**	−0.076***
		(0.025)	(0.031)	(0.034)	(0.029)
same_ins		−0.150***	−0.214***	−0.202	−0.204
		(0.044)	(0.067)	(0.146)	(0.128)
g2v		−0.084**	−0.065***	−0.042*	−0.056**
		(0.035)	(0.007)	(0.022)	(0.024)
v2g		−0.074***	−0.109***	−0.092***	−0.129***
		(0.007)	(0.013)	(0.017)	(0.020)
comp1		−0.254***	−0.341***	−0.338**	−0.309***
		(0.080)	(0.013)	(0.147)	(0.102)
comp2		−0.147**	−0.224***	−0.200***	−0.177***
		(0.066)	(0.005)	(0.023)	(0.019)
base param2				0.024**	0.005
				(0.012)	(0.011)
comp2 param1				−0.020	−0.070**
				(0.037)	(0.030)
comp1 param2				−0.0004	−0.039
				(0.148)	(0.107)
comp1 param1				0.017	−0.022
				(0.149)	(0.101)
vcg:res_price			0.019	0.002	0.018
			(0.046)	(0.045)	(0.046)
vcg:same_ins			0.133*	0.157	0.133*
			(0.069)	(0.146)	(0.069)
vcg:g2v			−0.032	−0.054	−0.035
			(0.058)	(0.062)	(0.049)
vcg:v2g			0.077***	0.065***	0.055***
			(0.017)	(0.023)	(0.012)
vcg:comp1			0.213***	0.276*	0.211***
			(0.022)	(0.150)	(0.021)
vcg:comp2			0.159***	0.142***	0.159***
			(0.005)	(0.026)	(0.005)
vcg:base param2				−0.030**	
				(0.014)	
vcg:as.factor(max_val)30800				0.005	
				(0.040)	
vcg:comp1 param2				−0.067	
				(0.151)	
vcg:comp1 param1				−0.087	
				(0.151)	
res_price:comp2 param1					0.017
					(0.036)
same_ins:comp2 param1					0.077
					(0.071)
g2v:comp2 param1					0.059
					(0.043)
v2g:comp2 param1					0.147***
					(0.037)
same_ins:comp1 param2					0.025
					(0.053)
Constant	0.740***	0.884***	0.869***	0.858***	0.867***
	(0.051)	(0.006)	(0.005)	(0.001)	(0.004)

Note:	*p<0.1; **p<0.05; ***p<0.01

28

Table 8: Revenue; reference vars are *GSP* and *Baseline.* Clustered at the session level

| | *Revenue:* | | |
	(1)	(2)	(3)
vcg	−0.069		−0.142
	(0.071)		(0.140)
res_price		0.026	−0.051
		(0.088)	(0.081)
same_ins		−0.081	−0.177*
		(0.093)	(0.097)
g2v		−0.055	−0.128
		(0.088)	(0.080)
v2g		−0.056	−0.145*
		(0.089)	(0.082)
comp1		−0.105	−0.176**
		(0.088)	(0.081)
comp2		−0.112	−0.251***
		(0.095)	(0.080)
vcg:res_price			0.115
			(0.140)
vcg:same_ins			0.153
			(0.152)
vcg:g2v			0.106
			(0.141)
vcg:v2g			0.138
			(0.140)
vcg:comp1			0.117
			(0.140)
vcg:comp2			0.238*
			(0.140)
Constant	0.989***	0.990***	1.081***
	(0.051)	(0.088)	(0.080)
Note:		*p<0.1; **p<0.05; ***p<0.01	

Table 9: Revenue – reference vars *GSP*, *baseline*, *base param1*, *comp1 param1*. Clustered at the session level

			Dependent variable:		
	(1)	(2)	(3)	(4)	(5)
vcg	−0.069		−0.142	−0.085	−0.141
	(0.071)		(0.140)	(0.111)	(0.138)
res_price		0.026	−0.051	0.002	−0.082
		(0.088)	(0.081)	(0.122)	(0.095)
same_ins		−0.081	−0.177*	−0.159**	−0.240***
		(0.093)	(0.097)	(0.066)	(0.082)
g2v		−0.055	−0.128	−0.074	−0.157
		(0.088)	(0.080)	(0.122)	(0.095)
v2g		−0.056	−0.145*	−0.095	−0.186**
		(0.089)	(0.082)	(0.123)	(0.095)
comp1		−0.105	−0.176**	−0.192***	−0.252***
		(0.088)	(0.081)	(0.065)	(0.074)
comp2		−0.112	−0.251***	−0.195	−0.278***
		(0.095)	(0.080)	(0.122)	(0.095)
base param2				−0.066**	−0.140***
				(0.031)	(0.044)
comp2 param2				−0.077	−0.041
				(0.105)	(0.061)
comp2 param1				−0.092	−0.038
				(0.102)	(0.061)
comp1 param2				−0.030*	0.020
				(0.016)	(0.021)
vcg:res_price			0.115	−0.049	0.114
			(0.140)	(0.160)	(0.138)
vcg:same_ins			0.153	0.025	0.151
			(0.152)	(0.122)	(0.149)
vcg:g2v			0.106	−0.061	0.104
			(0.141)	(0.161)	(0.138)
vcg:v2g			0.138	−0.023	0.141
			(0.140)	(0.161)	(0.138)
vcg:comp1			0.117	0.025	0.116
			(0.140)	(0.111)	(0.138)
vcg:comp2			0.238*	0.074	0.237*
			(0.140)	(0.160)	(0.138)
vcg:base param2				−0.116**	
				(0.056)	
vcg:comp2 param2				0.104	
				(0.117)	
vcg:comp2 param1				0.110	
				(0.114)	
vcg:comp1 param2				0.072***	
				(0.017)	
res_price:comp2 param2					0.007
					(0.006)
same_ins:comp2 param2					0.038***
					(0.013)
g2v:comp2 param2					0.004
					(0.018)
v2g:comp2 param2					0.021
					(0.019)
same_ins:comp1 param2					0.007
					(0.030)
Constant	0.989***	0.990***	1.081***	1.112***	1.147***
	(0.051)	(0.088)	(0.080)	(0.065)	(0.073)
Note:					*p<0.1; **p<0.05; ***p<0.01

Figure 12: Subject User Interface
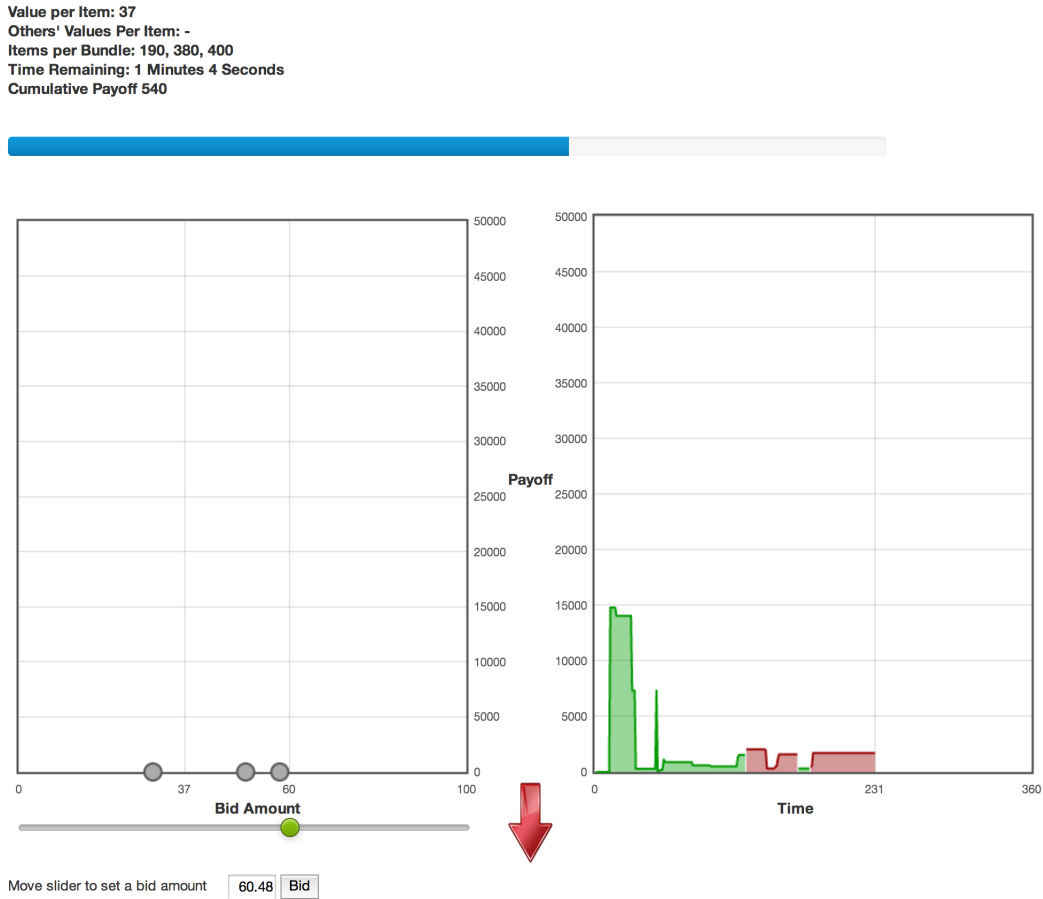
References

[1] Athey, S., D. Nekipelov, "A Structural Model of Sponsored Search Advertising Auctions" *Journal of Political Economy*, forthcoming.

[2] Bailey, Michael Seminar presentation, 2015, *Stanford GSB Mechanism Design*

[3] Börgers, T., I. Cox, M. Pesendorfer, and V. Petricek, "Equilibrium Bids in Sponsored Search Auctions: Theory and Evidence" *American Economic Journal: Microeconomics*, 5(4), 2013, 163-87.

[4] Che, Y., S. Choi, and J. Kim, *"An Experimental Study of Sponsored-Search Auctions"* March, 2013. Available at http://www.homepages.ucl.ac.uk/ uctpsc0/Research/CCK.pdf

[5] Edelman, B., M. Ostrovsky, and M. Schwarz, "Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords" *American Economic Review*, 97(1), 2007, 242-259.

[6] Friedman, D. and R. Oprea, "A Continuous Dilemma," *Am. Econ. Review*, 102(1), 2012, 337-363.

[7] Fukuda, E., Y. Kamijo, A. Takeuchi, M. Masui, Y. Funaki, "Theoretical and Experimental Investigation of Performance of Keyword Auction Mechanisms" *RAND Journal of Economics*, 44(3), Fall 2013, 438-461.

[8] Goldman, M. and J. Rao, "Experiments as Instruments: Heterogeneous Position Effects in Sponsored Search Auctions" *Unpublished Job Market Paper*, November, 2014.

[9] Hlavac, Marek (2014). stargazer: LaTeX code and ASCII text for well-formatted regression and summary statistics tables.

[10] Milgrom, P. "Assignment Messages and Exchanges" *American Economic Journal: Microeconomics*, 1(2), 2009, 95-113.

[11] Moldovanu, B. and A. Sela "The Optimal Allocation of Prizes in Contests" *American Economic Review*, 91(3), 2001, 542-58.

[12] Ember, Sydney "Digital Ad Spending Expected to Soon Surpass TV" *The New York Times*, Dec. 7, 2015

[13] Noti, G., N. Nisan, and I. Yaniv, "An Experimental Evaluation of Bidders' Behavior in Ad Auctions" *International World Wide Web Conference Committee*, WWW'14, April 7-11, 2014.

[14] Oprea, R., K. Henwood, and D. Friedman, "Separating the Hawks from the Doves: Evidence from continuous time laboratory games" *Journal of Economic Theory*, 146(6), 2011, 2206-2225.

[15] Ostrovsky, M and M. Schwarz, "Reserve Prices in Internet Advertising Auctions: A Field Experiment" *Research Papers 2054*, Stanford University, Graduate School of Business, 2009.

[16] Software for Continuous Game Experiments, James Pettit, Daniel Friedman, Curtis Kephart, and Ryan Oprea, Experimental Economics 17:4 pp. 631-648 (2014).

[17] Roth, A. and M. Sotomayor "Two-Sided Matching" *Cambridge University Press*, 1990.

[18] Simon, L. K. and M. B. Stinchcombe, "Continuous Time: Pure Strategies" *Econometrica*, 57(5), 1989, 1171-1214.

[19] Varian, H. R. and C. Harris, "The VCG Auction in Theory and Practice" *American Economic Review*, 104(5), 2014, 442-45.

[20] Varian, H. R., "The Economics of Internet Search" *'Angelo Costa' Lecture Serie*, SIPI Spa, issue Lect. VII.

[21] Varian, H. R., "Online Ad Auctions" *American Economic Review*, 99(2), 2009, 430-34.

[22] Varian, H. R., "Position Auctions" *Int. J. of Industrial Organization*, 25(1), 2007, 1163-1178.