MS&E 233
Game Theory, Data Science and AI
Lecture 10

Vasilis Syrgkanis

Assistant Professor

Management Science and Engineering

(by courtesy) Computer Science and Electrical Engineering

Institute for Computational and Mathematical Engineering

Computational Game Theory for Complex Games

1
- Basics of game theory and zero-sum games (T)
- Basics of online learning theory (T)
- Solving zero-sum games via online learning (T)
- *HW1: implement simple algorithms to solve zero-sum games*
- Applications to ML and AI (T+A)
- *HW2: implement boosting as solving a zero-sum game*

2
- Basics of extensive-form games
- Solving extensive-form games via online learning (T)
- *HW3: implement agents to solve very simple variants of poker*

3
- General games, equilibria and online learning (T)
- Online learning in general games
- *HW4: implement no-regret algorithms that converge to correlated equilibria in general games*

Data Science for Auctions and Mechanisms

4
- Basics and applications of auction theory (T+A)
- **Basic Auctions and Learning to bid in auctions (T)**
- *HW5: implement bandit algorithms to bid in ad auctions*

5
- Optimal auctions and mechanisms (T)
- Simple vs optimal mechanisms (T)
- *HW6: calculate equilibria in simple auctions, implement simple and optimal auctions, analyze revenue empirically*

6
- Optimizing mechanisms from samples (T)
- Online optimization of auctions and mechanisms (T)
- *HW7: implement procedures to learn approximately optimal auctions from historical samples and in an online manner*

Further Topics

7
- Econometrics in games and auctions (T+A)
- A/B testing in markets (T+A)
- *HW8: implement procedure to estimate values from bids in an auction, empirically analyze inaccuracy of A/B tests in markets*

Guest Lectures

- Mechanism Design for LLMs, Renato Paes Leme, Google Research
- Auto-bidding in Sponsored Search Auctions, Kshipra Bhawalkar, Google Research

Sum: Auction Applications

- Traditionally, selling of luxury goods, art
- Digital auction markets for goods (eBay)
- Energy markets
- Digital ad markets (sponsored search, display ads, amazon ads)
- Spectrum auctions
- Government procurement auctions
- Web3.0 transaction protocols

Sum: First Price

- First Price is arguably the simplest auction rule
- It can be hard to strategize in such an auction
- The auction can lead to inefficient allocations

- Though approximately efficient
- Still used in practice in many settings (e.g. online advertising, government procurement)
- Primarily because it has very transparent rules

Sum: Second Price

- Second Price is arguably the simplest truthful auction rule
- It is very easy to strategize in such an auction (be truthful)
- Auction always leads to efficient allocations (highest value wins)
- Auction can be run very quickly (computationally efficient)

- Still not always the auction used in many places
- Primarily because it has not very transparent rules
- Susceptible to collusion and manipulations by the auctioneer

Sponsored Search Auctions

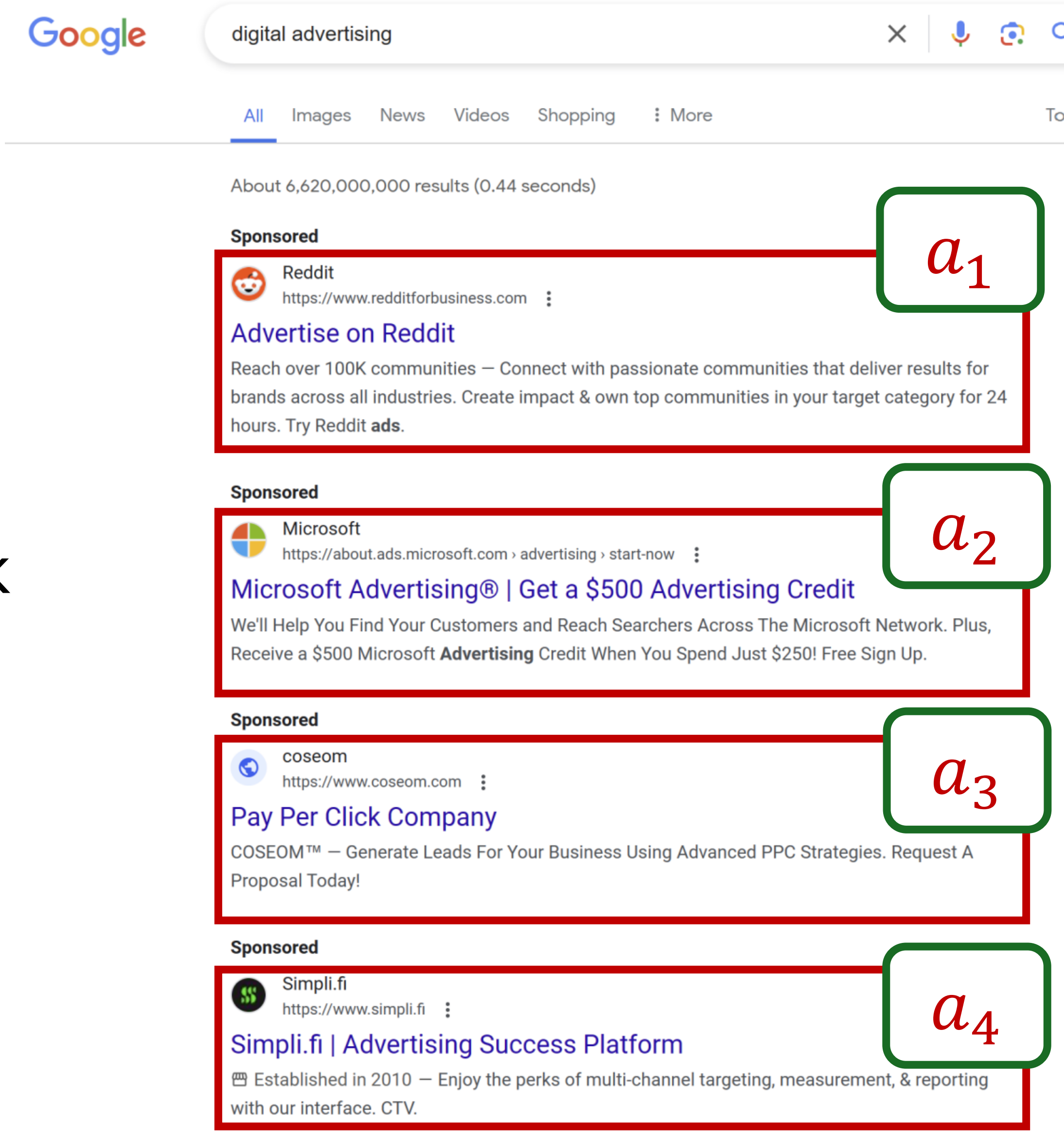
Sponsored Search Auctions

- Now we have many items to sell
- Slots on a web impressions

- Higher slots get more clicks!
- Each slot has some probability of click

$$a_1 > a_2 > \cdots > a_m$$

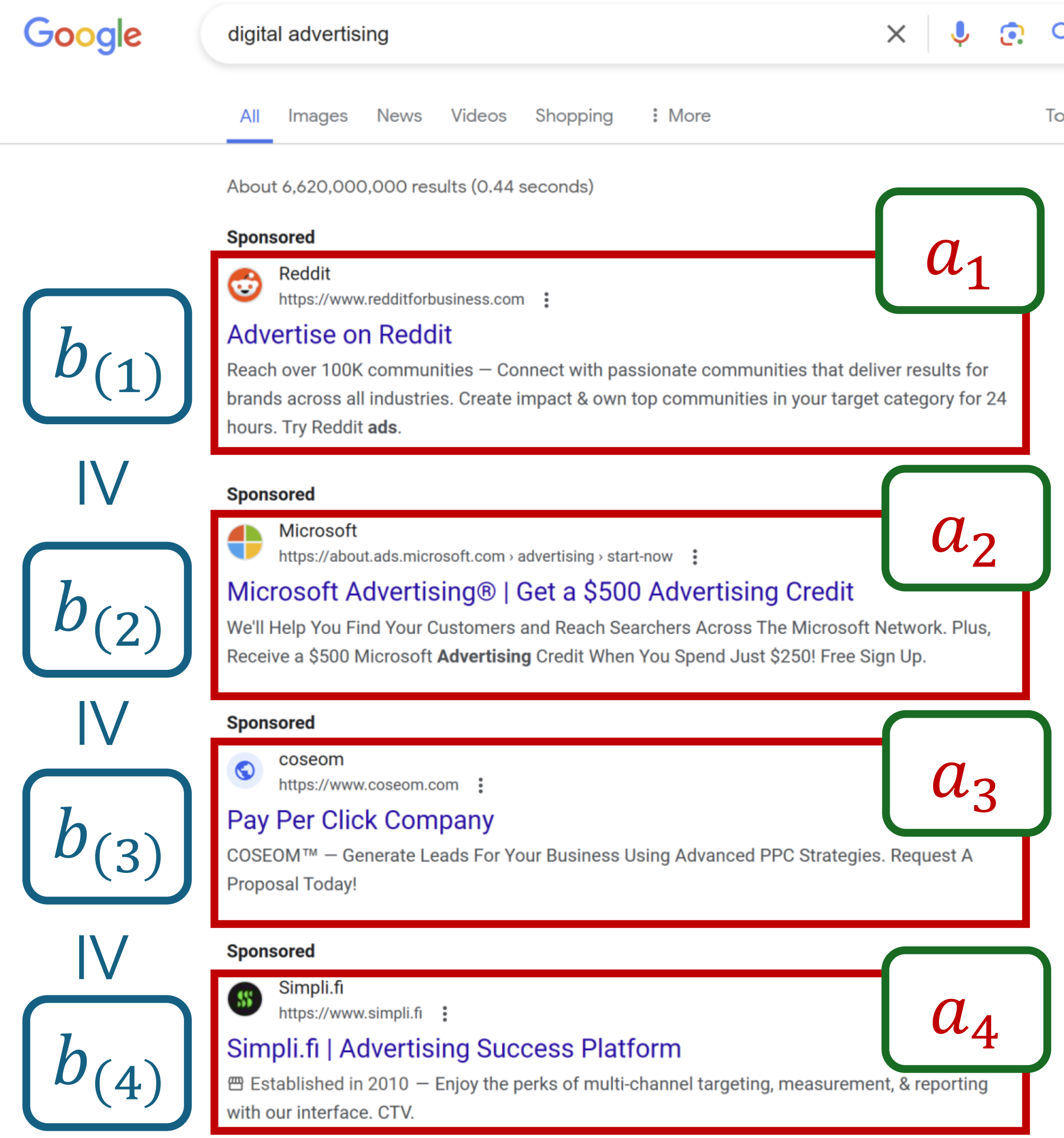
- Bidders have a value-per-click v_i



Generalized First Price (GFP) Auction

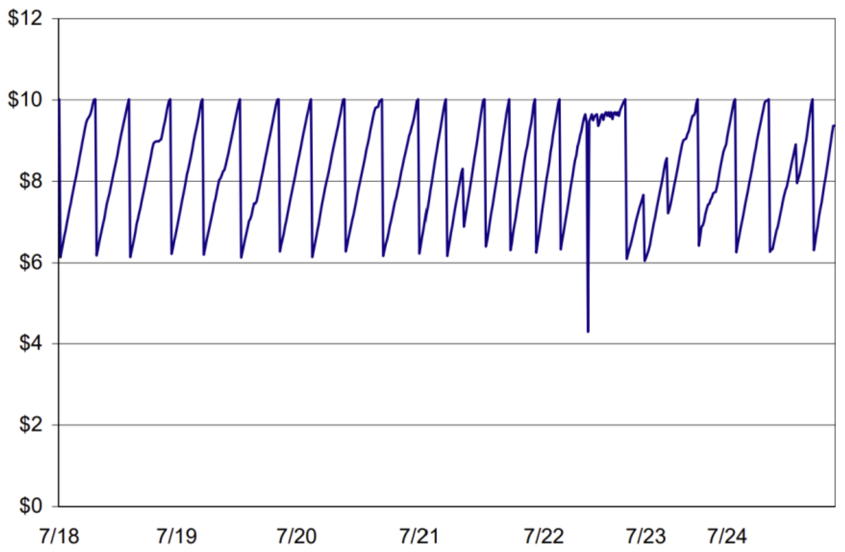
- Bidders submit a bid-per-click b_i
- Slots allocated in decreasing order of bids
- Bidder i is allocated slot $j_i(b)$
- Bidder pays their bid when clicked

$$u_i(b; v_i) = a_{j_i(b)} \cdot (v_i - b_i)$$

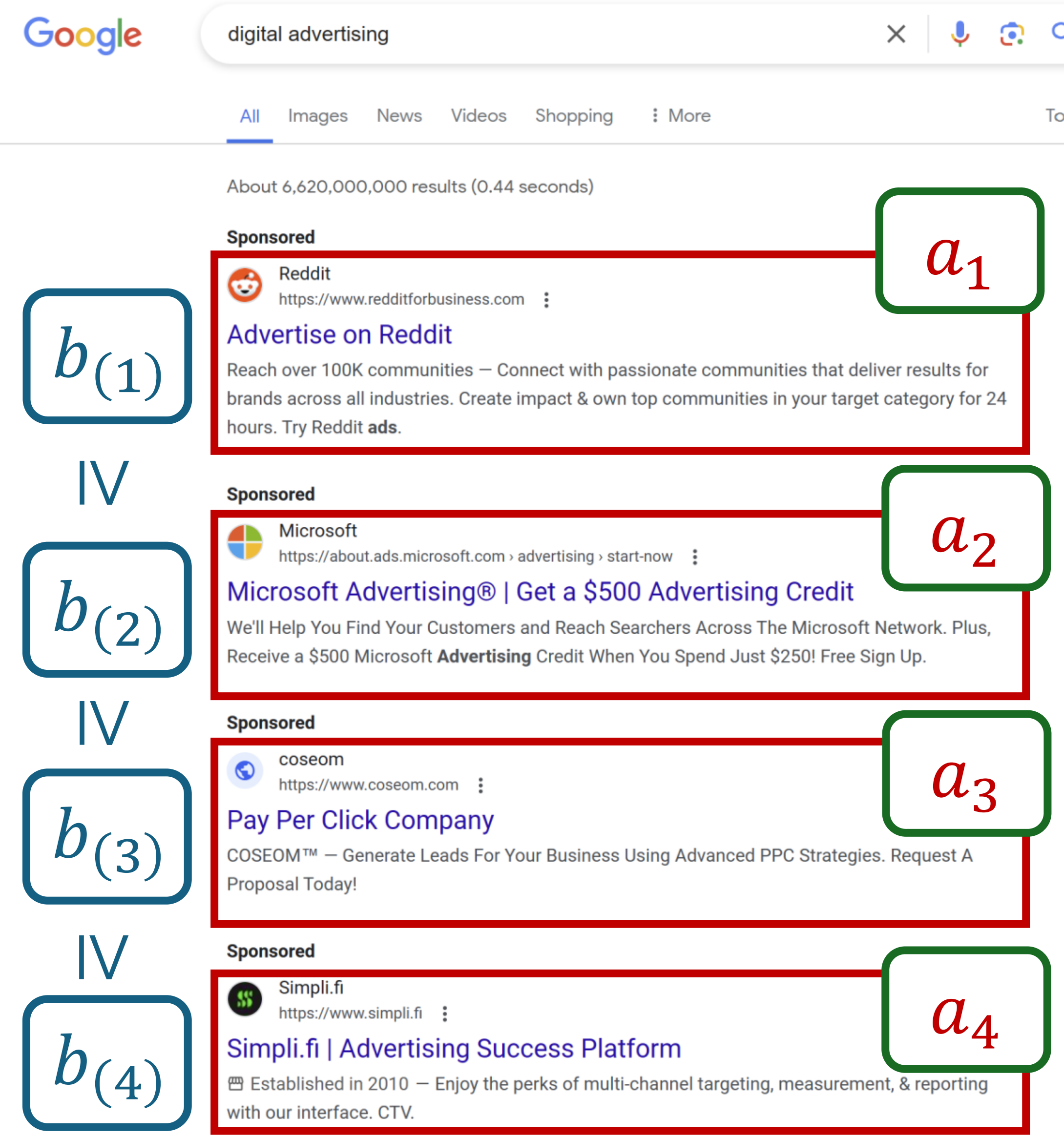


Generalized First Price (GFP) Auction

- The first auction that was used by Overture in late 90s
- Lead to weird bidding patterns



(b) 1 week

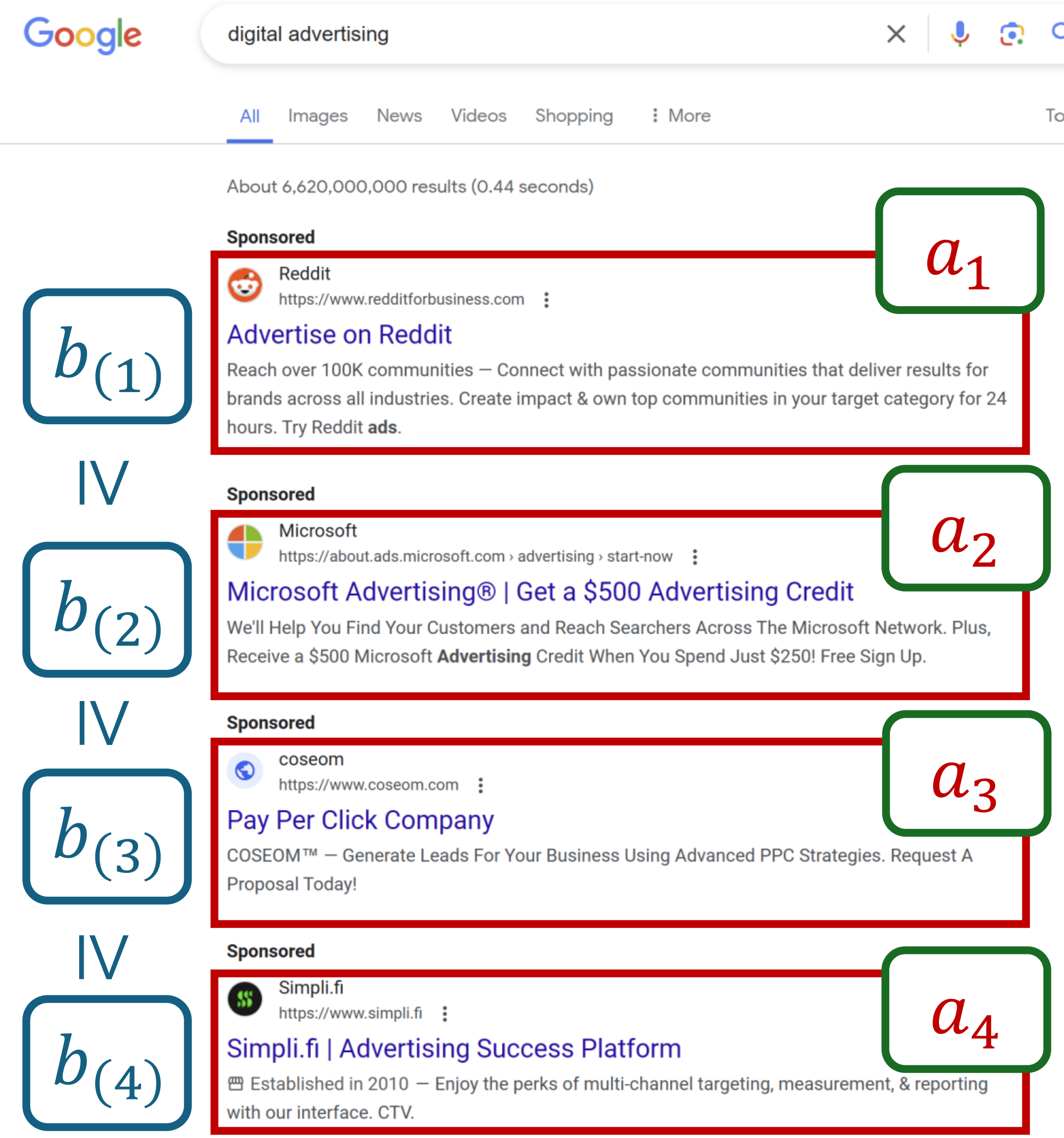


Credits: https://www.benedelman.org/publications/cycling-060703.pdf

Generalized Second Price (GSP) Auction

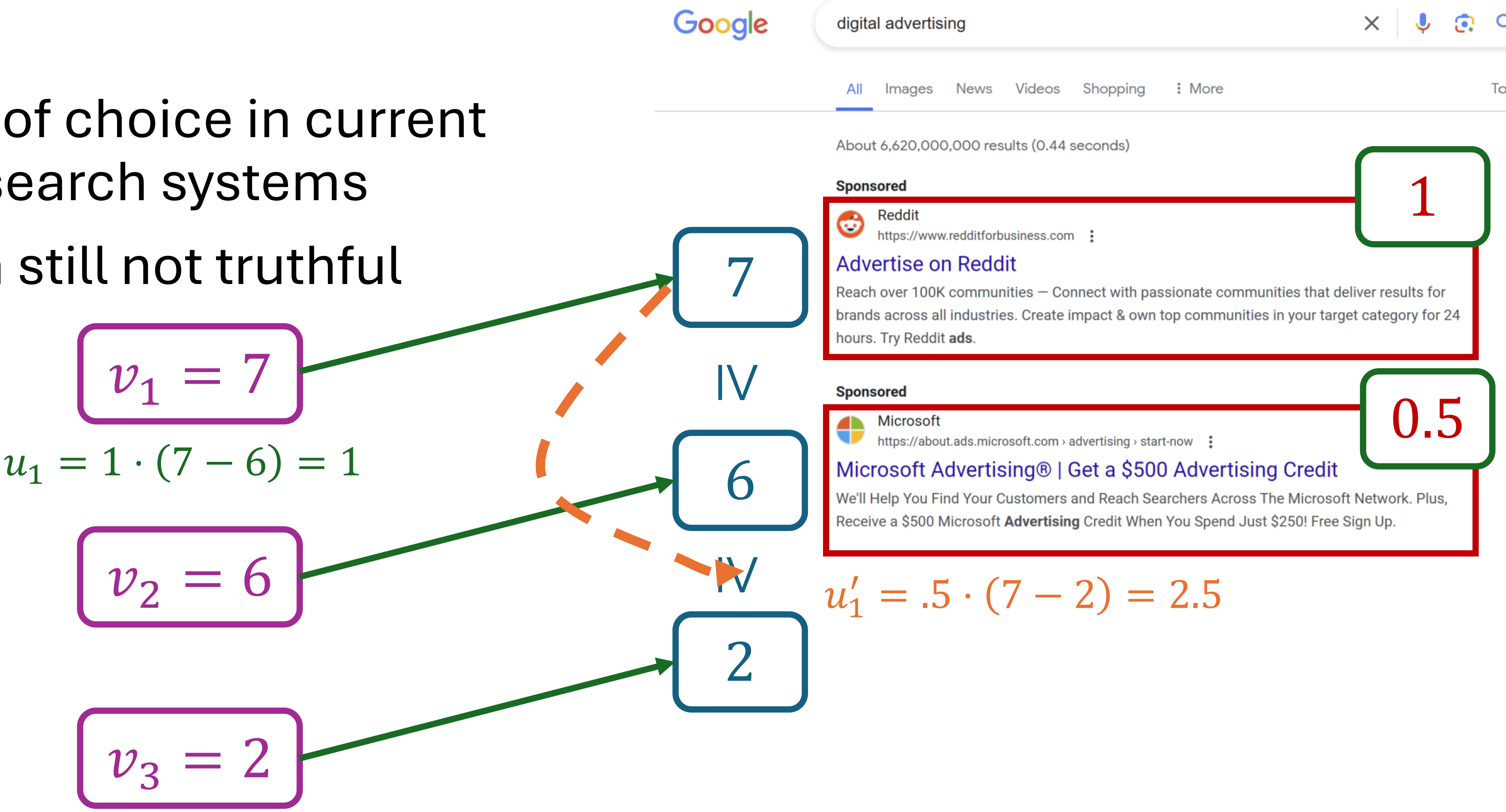
- Bidders submit a bid-per-click b_i
- Slots allocated in decreasing order of bids
- Bidder i is allocated slot $j_i(b)$
- Bidder pays the next highest bid when clicked

$$u_i(b; v_i) = a_{j_i(b)} \cdot \left(v_i - b_{(j_i(b)+1)}\right)$$



Generalized Second Price (GSP) Auction

- The auction of choice in current sponsored search systems
- Even though still not truthful



How would you turn GSP truthful?

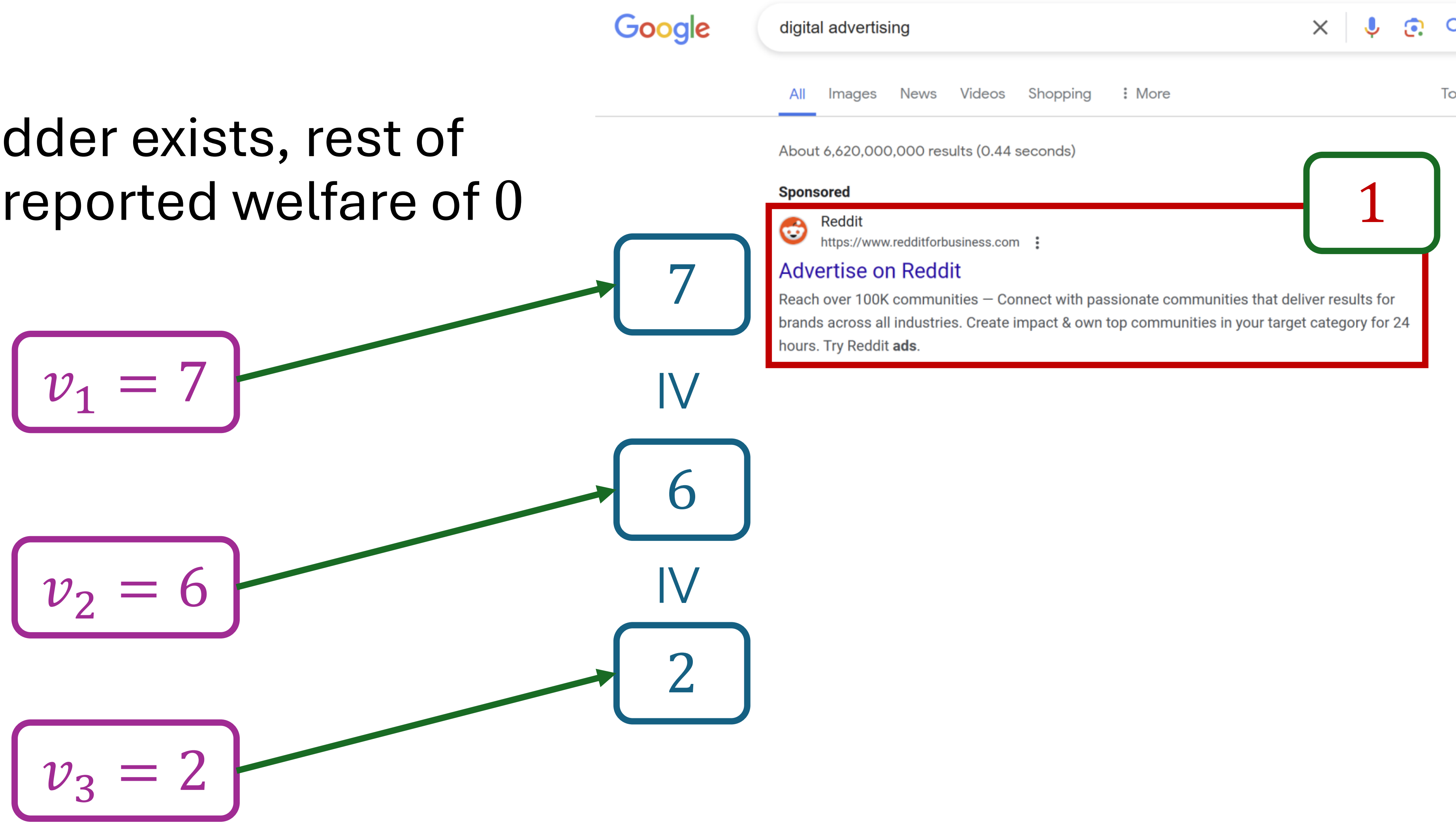
Right intuition, why Second-Price is truthful

- Second price is truthful **not because** we charge next highest bid
- Second price is truthful **not because** we charge smallest bid to maintain the same allocation

- Second price is truthful **because** we charged the winner their “externalities to the rest of society”

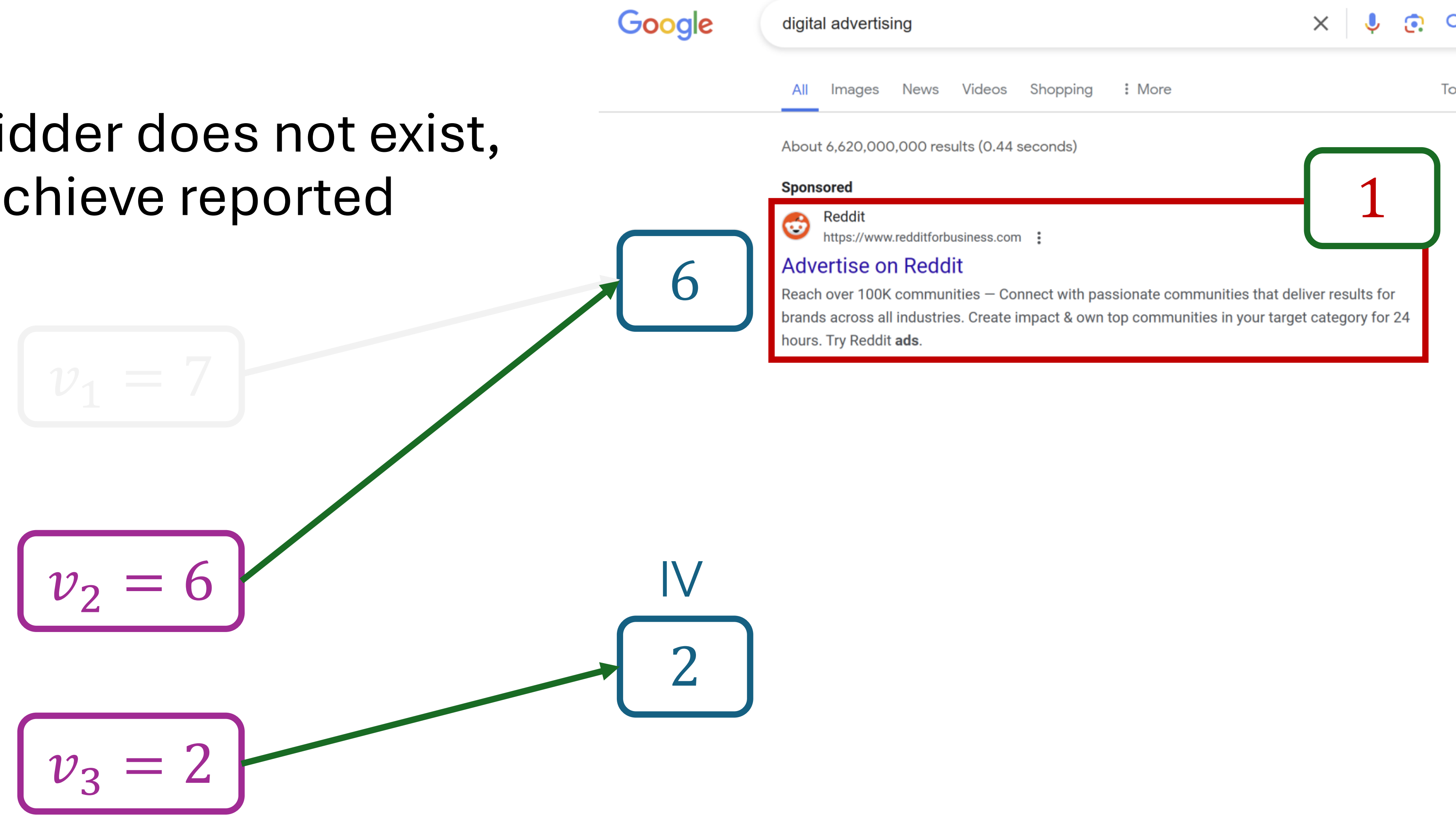
The Deep Reason why SP is Truthful

- When highest bidder exists, rest of players achieve reported welfare of 0



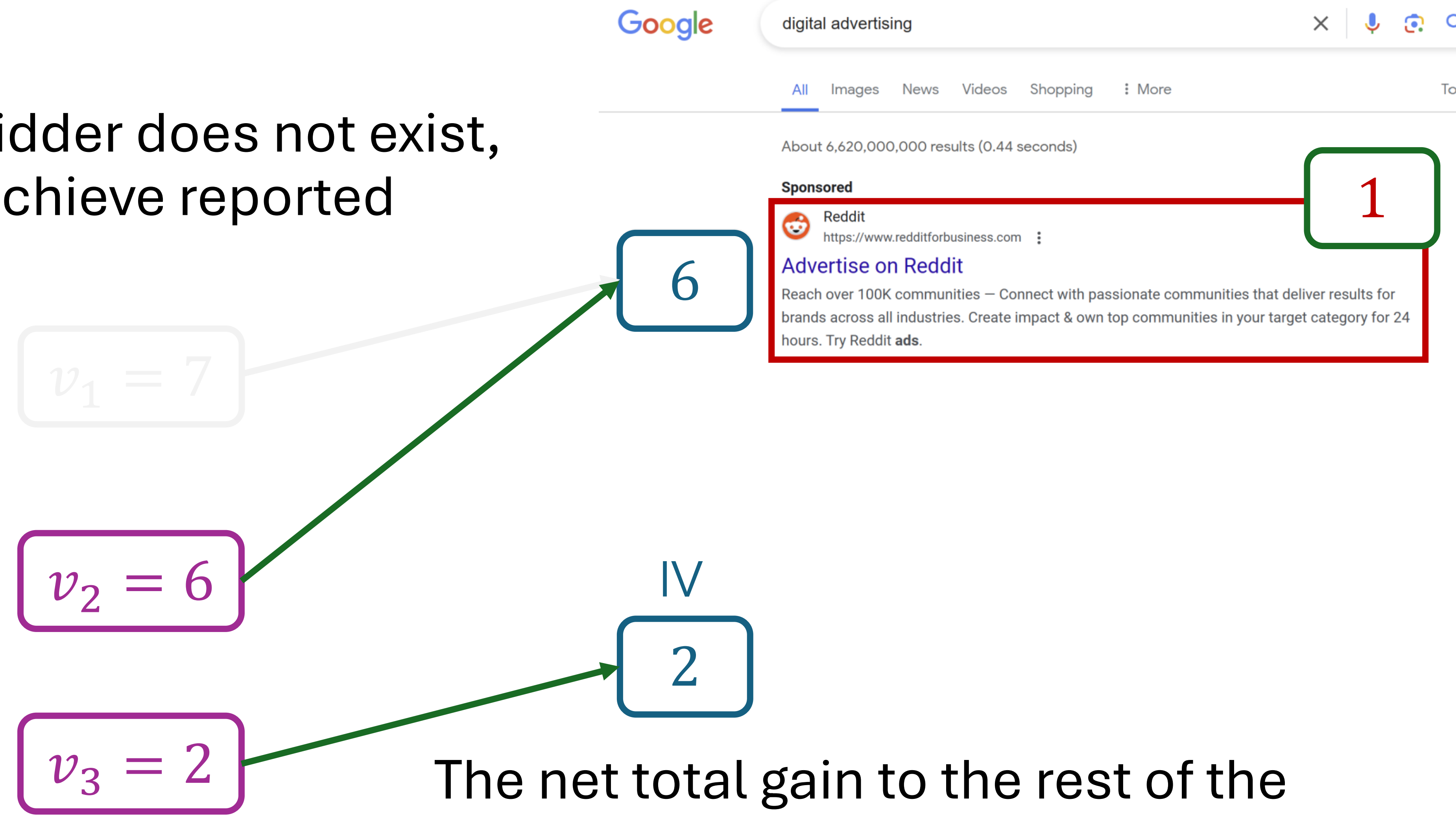
The Deep Reason why SP is Truthful

- When highest bidder does not exist, rest of players achieve reported welfare of 6



The Deep Reason why SP is Truthful

- When highest bidder does not exist, rest of players achieve reported welfare of 6



The net total gain to the rest of the bidders, from bidder 1 vanishing is 6

Right intuition, why Second-Price is truthful

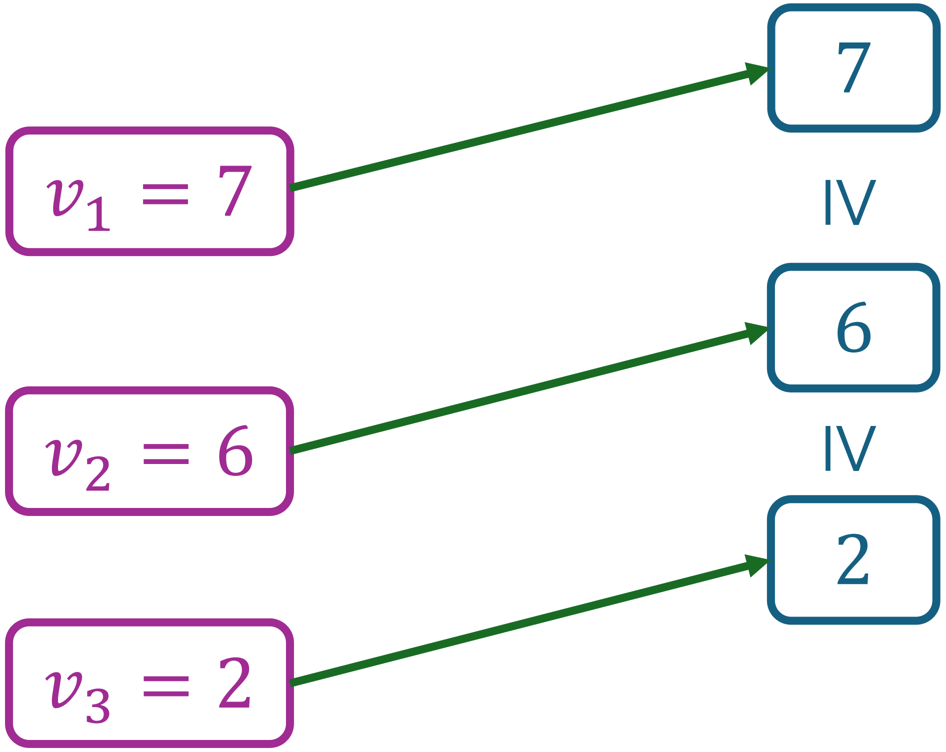
- Second price is truthful **because** we charged the winner their "externalities to the rest of society"

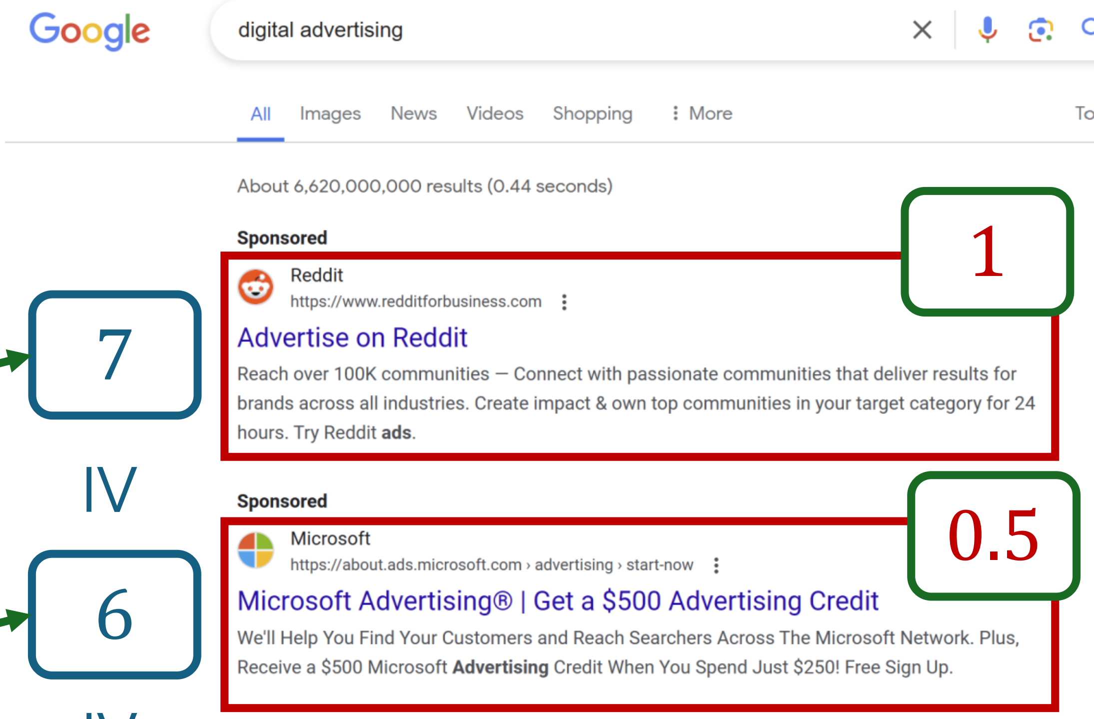
- When highest bidder exists, rest of players achieve reported welfare 0
- When highest bidder vanishes, rest of players achieve reported welfare
$$b_{(2)} = \text{second highest bid}$$
- The net total gain to the rest of the bidders, from bidder 1 vanishing is
$$b_{(2)} = \text{second highest bid}$$

- That's what we should charge the winner!

Let’s repeat this exercise with two slots

- When highest bidder exists, rest of players achieve reported welfare of ...?

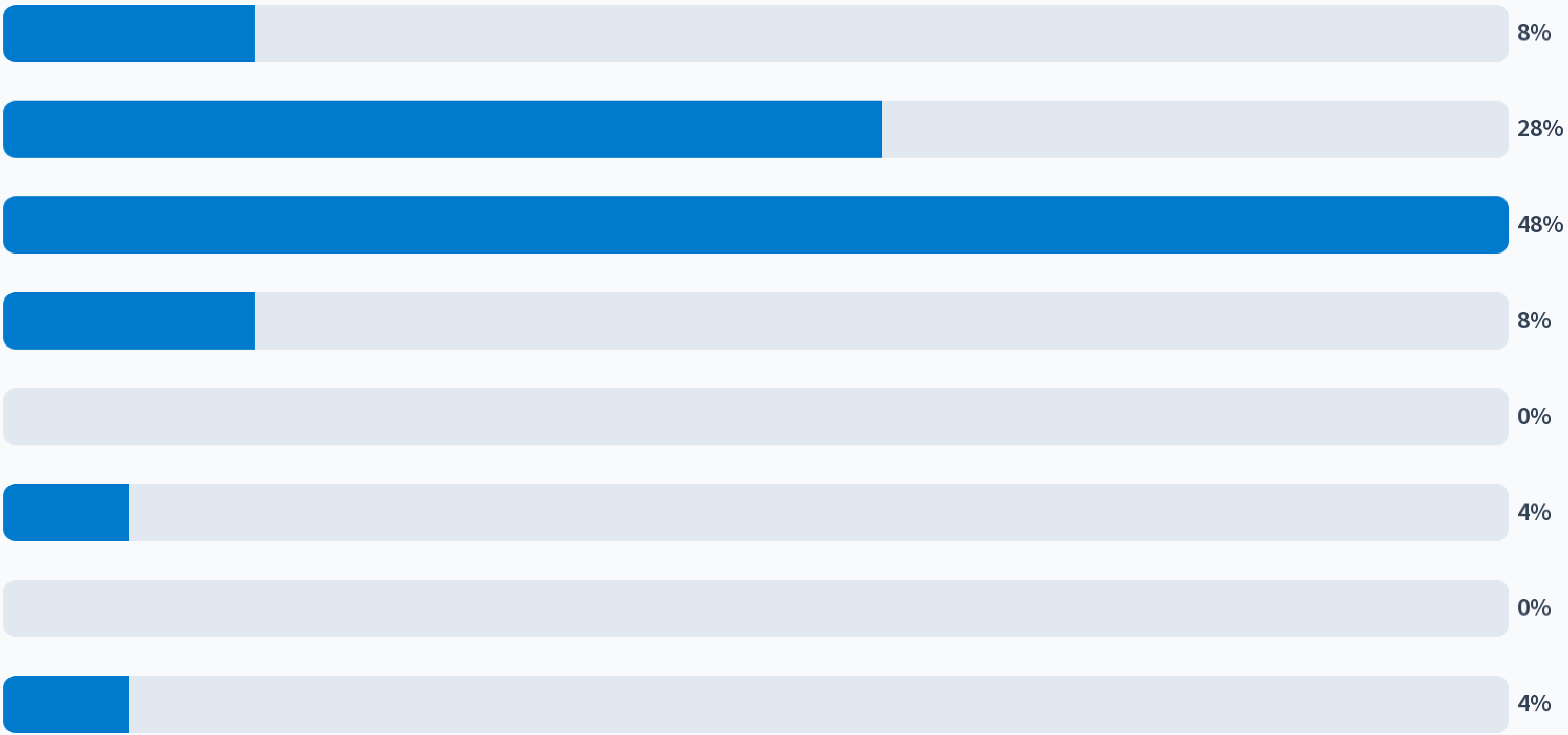




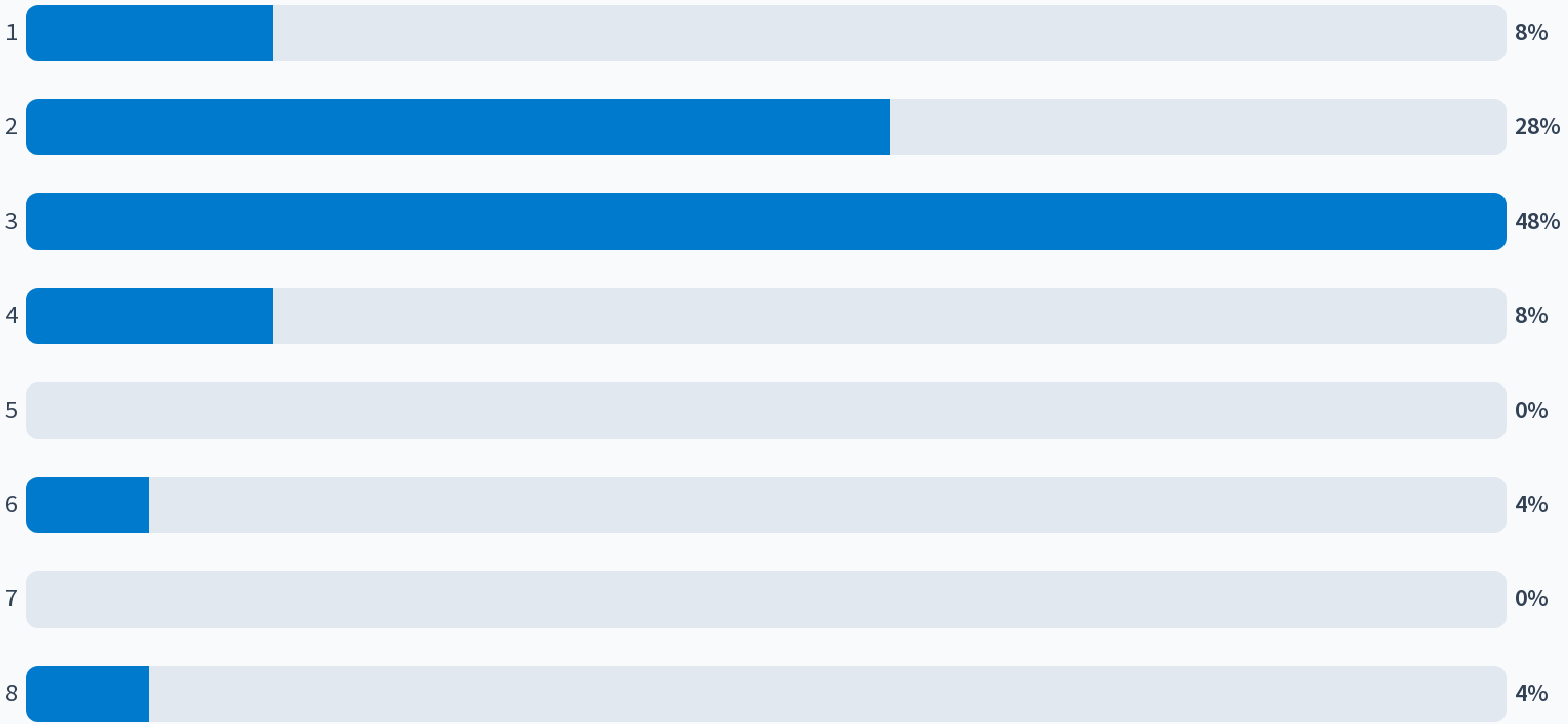
When the highest value bidder exists the rest of the players get a reported welfare of

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8

When the highest value bidder exists the rest of the players get a reported welfare of

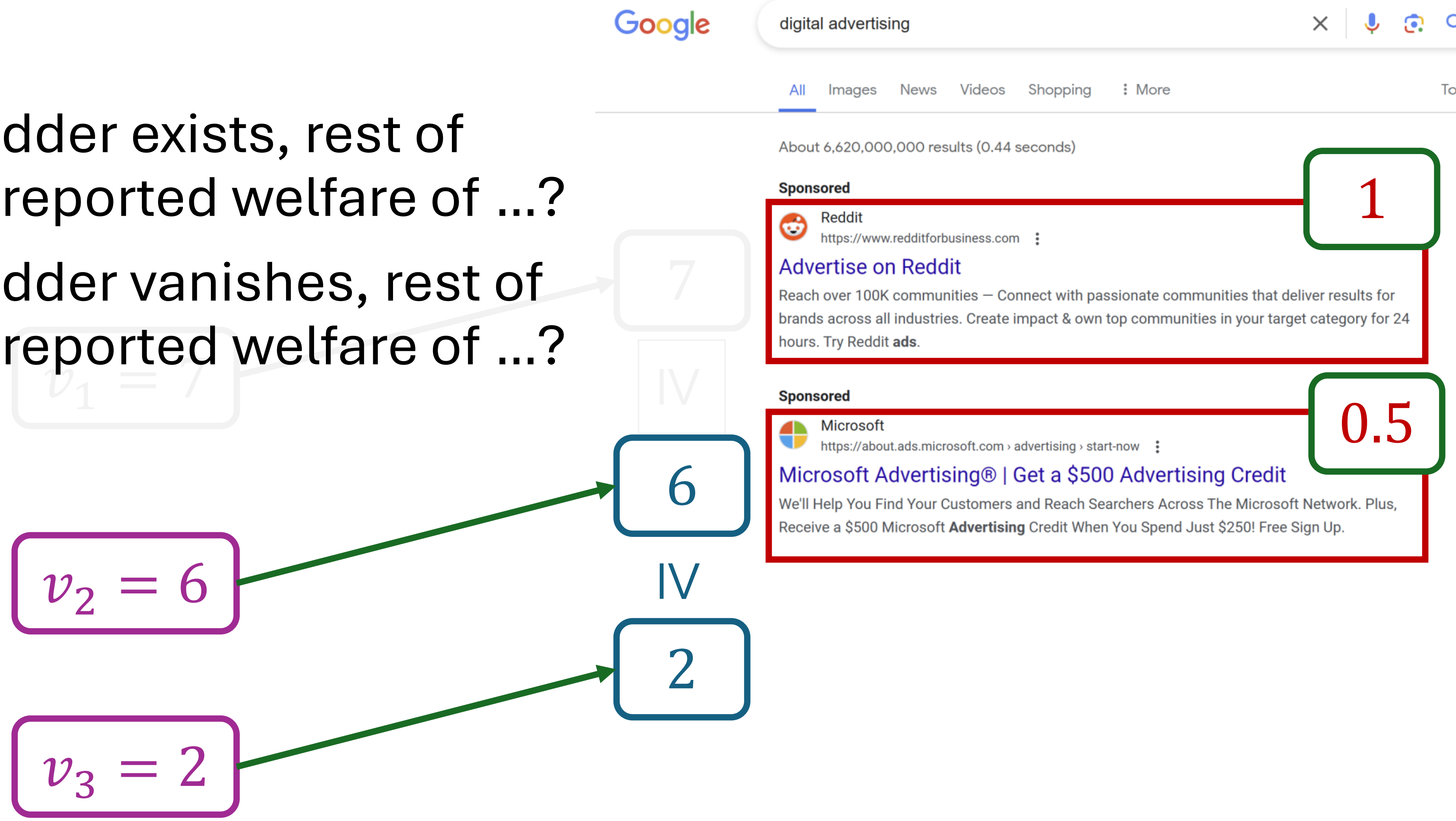


When the highest value bidder exists the rest of the players get a reported welfare of



Let’s repeat this exercise with two slots

- When highest bidder exists, rest of players achieve reported welfare of …?
- When highest bidder vanishes, rest of players achieve reported welfare of …?



When the highest value bidder vanishes the rest of the players get a reported welfare of

1

2

3

4

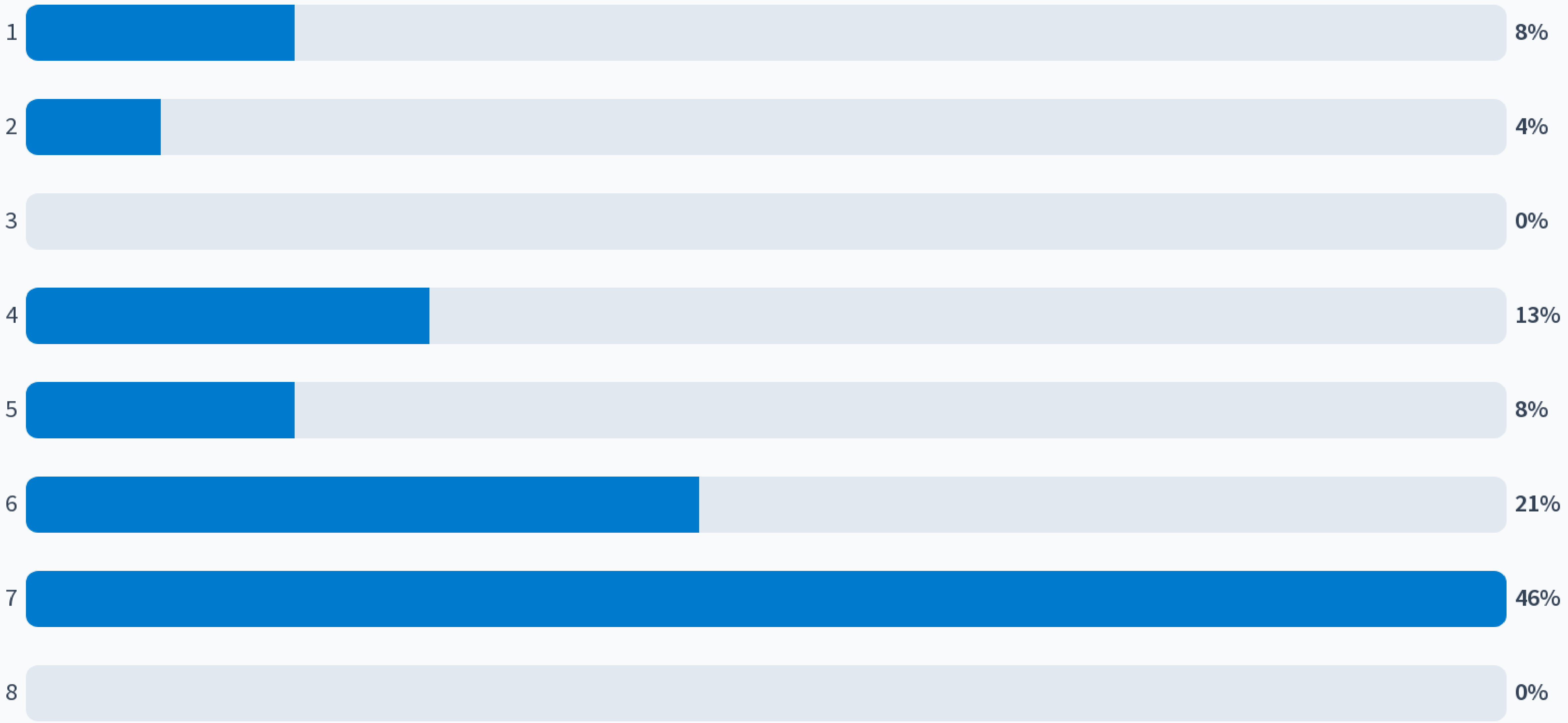
5

6

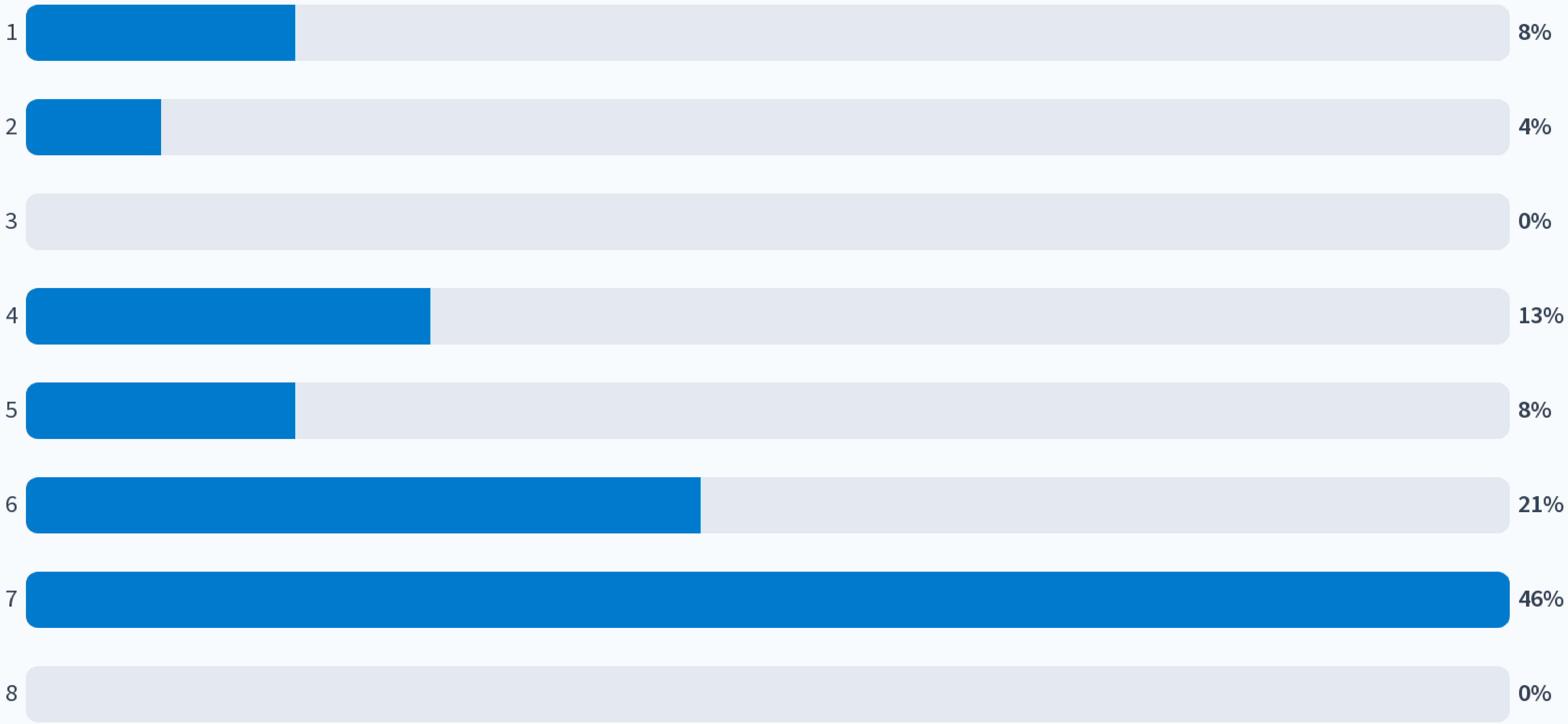
7

8

When the highest value bidder vanishes the rest of the players get a reported welfare of

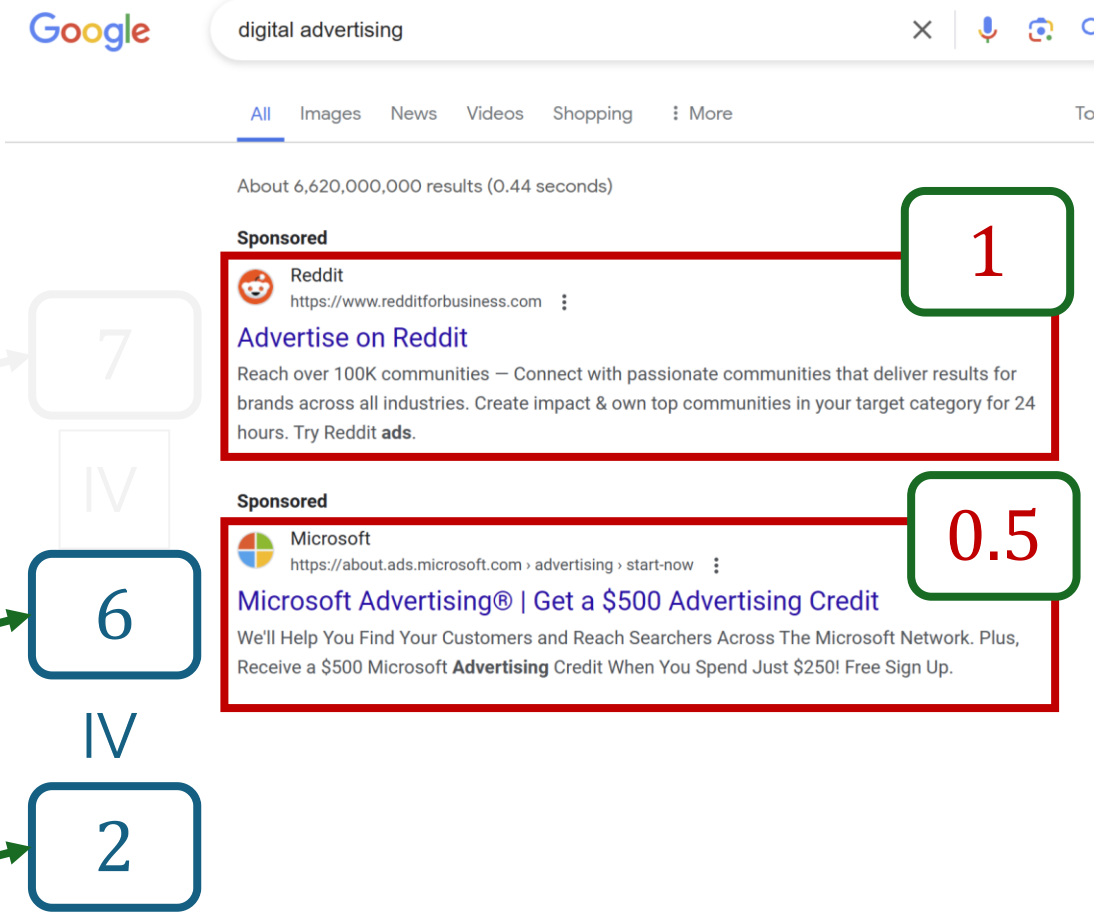


When the highest value bidder vanishes the rest of the players get a reported welfare of



Let's repeat this exercise with two slots

- When highest bidder exists, rest of players achieve reported welfare of ...?
- When highest bidder vanishes, rest of players achieve reported welfare of ...?



The net total gain to the rest of the bidders, from bidder 1 vanishing is ...

What about the second highest bidder?

- When second highest bidder exists, rest of players achieve reported welfare of 7

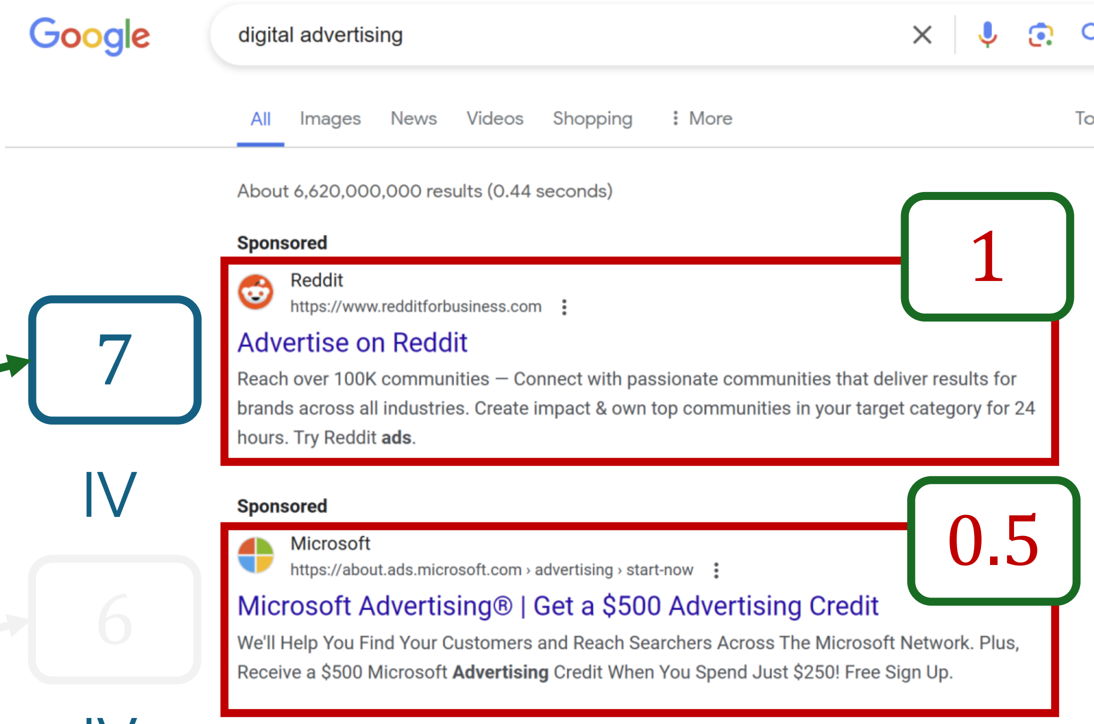
$v_1 = 7$ → 7

$\geq$

$\geq$

- When second highest bidder vanishes, rest of players achieve reported welfare of $7 + 1$

$v_3 = 2$ → 2



I should charge a total price of 1 (equivalently a price-per-bid of 2)

Bidders now don't have incentive to deviate

- The auction of choice in current sponsored search systems
- Even though still not truthful

$v_1 = 7$

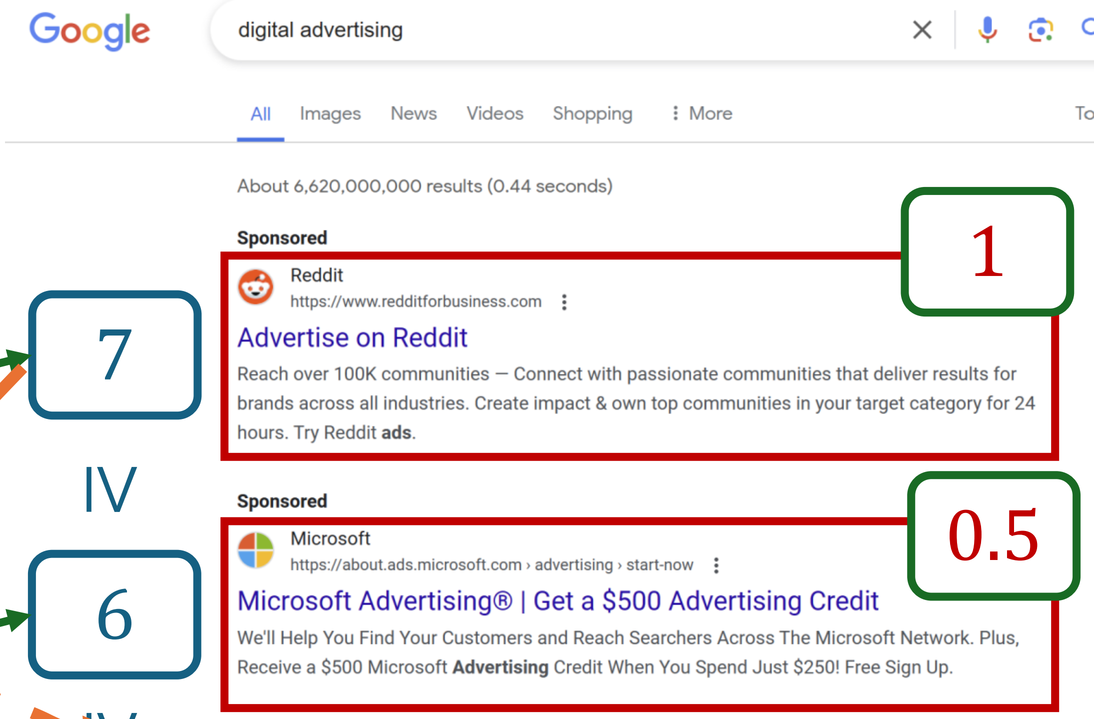
$u_1 = 1 \cdot (7 - 4) = 3$

$v_2 = 6$

$u_2 = .5 \cdot (6 - 2) = 2.5$

$v_3 = 2$

$7 \geq 6 \geq 2$



1

0.5

$u_1' = .5 \cdot (7 - 2) = 2.5$

How much utility do bidders receive?

$$\textbf{Externality} = \text{RWelfare of Others without me} - \overset{\text{Reported Welfare}}{\boxed{\text{RWelfare}}}\text{ of Others with me}$$

$$\textbf{Utility} = \text{Value of my Allocation} - \text{Payment}$$

How much utility do bidders receive?

Reported Welfare

$$\textbf{Externality} = \text{RWelfare of Others without me} - \boxed{\text{RWelfare}}\text{ of Others with me}$$

$$\textbf{Utility} = \text{Value of my Allocation} - \text{Payment}$$

If we set payment = externality

$$\text{Value of my Allocation} - \text{RWelfare of Others without me} + \text{RWelfare of Others with me}$$

How much utility do bidders receive?

$$\textbf{Externality} = \text{RWelfare of Others without me} - \overset{\text{Reported Welfare}}{\boxed{\text{RWelfare}}}\ \text{of Others with me}$$

$$\textbf{Utility} = \text{Value of my Allocation} - \text{Payment}$$

If we set payment = externality

$$\text{Value of my Allocation} - \text{RWelfare of Others without me} + \text{RWelfare of Others with me}$$

When I'm truthful:

$$\text{Value of my Allocation} + \text{RWelfare of Others with me} = \text{Total RWeflare with me}$$

How much utility do bidders receive?

$$\textbf{Externality} = \text{RWelfare of Others without me} - \overset{\text{Reported Welfare}}{\boxed{\text{RWelfare}}}\ \text{of Others with me}$$

$$\textbf{Utility} = \text{Value of my Allocation} - \text{Payment}$$

If we set payment = externality

$$\text{Value of my Allocation} - \text{RWelfare of Others without me} + \text{RWelfare of Others with me}$$

When I’m truthful:

$$\text{Value of my Allocation} + \text{RWelfare of Others with me} = \text{Total RWeflare with me}$$

When I’m truthful my utility is as simple as:

$$\textbf{Utility} = \text{Total RWeflare with me} - \text{Total RWelfare without me}$$

Can we ever charge bidders more than value?

- If we set payment = externality, and bidder is truthful

$$\textbf{Utility} = \text{Total RWeflare with me} - \text{Total RWelfare without me}$$

- If the auction always chooses the outcome that maximizes the reported welfare, then

$$\text{Total RWeflare with me} \geq \text{Total RWelfare without me}$$

Why is the mechanism truthful?

- If we set payment = externality, and bidder is truthful

$$\textbf{Utility} = \text{Total RWeflare with me} - \text{Total RWelfare without me}$$

- My bid does not affect the Total RWelfare without me!
- RWelfare only depends on the chosen allocation, not payments
- Trying to choose a bid b_i that leads to allocation x that maximizes

$$\text{Total RWeflare with me(x)}$$

Intuition: Why is the mechanism truthful?

- If we set payment = externality, and bidder is truthful

$$\textbf{Utility} = \text{Total RWeflare with me} - \text{Total RWelfare without me}$$

- My bid does not affect the Total RWelfare without me!
- RWelfare only depends on the chosen allocation, not payments

- If I’m truthful the auctioneer chooses the allocation that maximizes exactly this quantity and hence that maximizes my utility.

The Vickrey-Clarke-Groves (VCG) Mechanism

General Auction (Mechanism Design) Setting

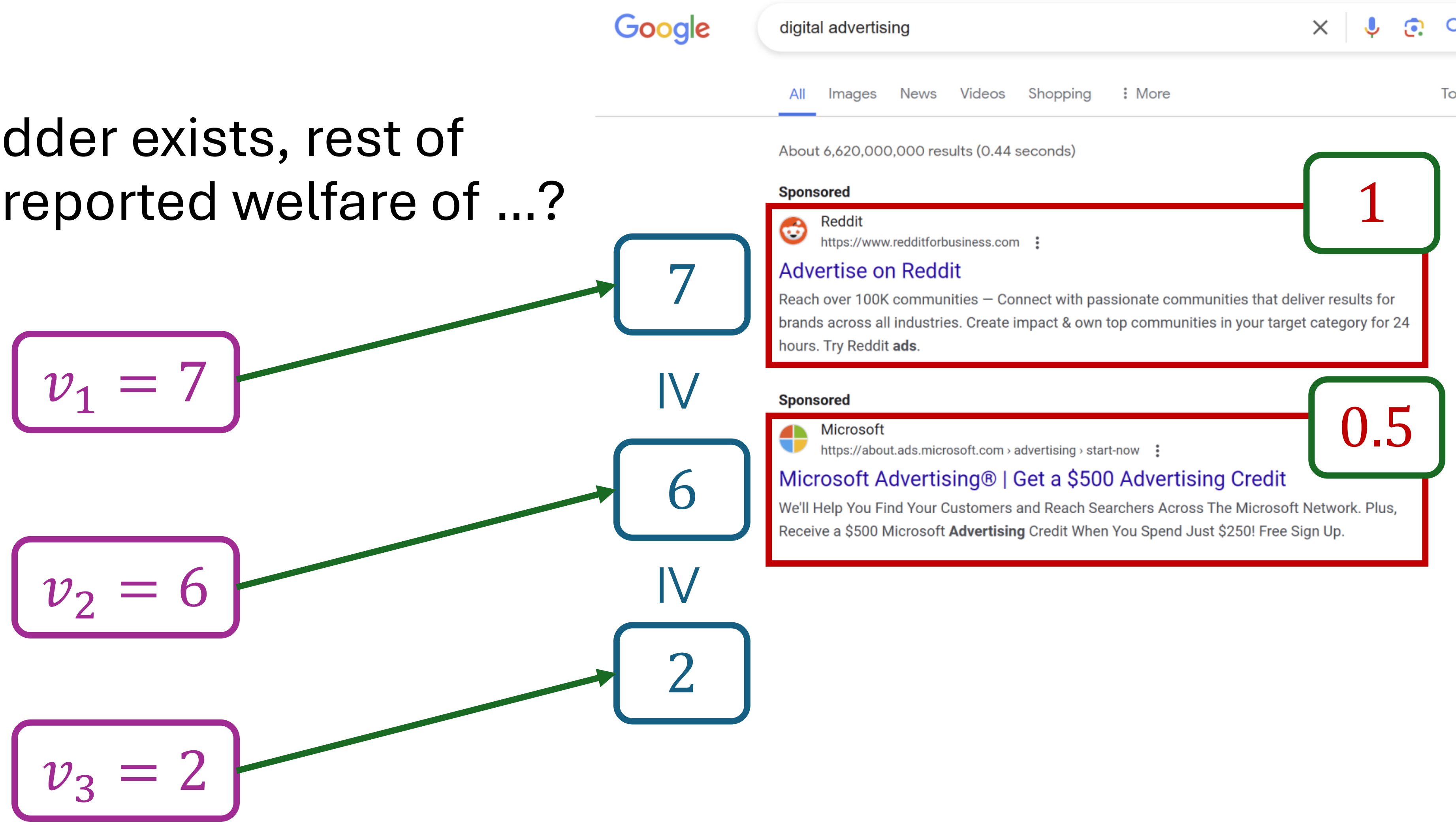
- Auctioneer (Designer) wants to choose among set of outcomes O
- Each bidder i has some value for each outcome $v_i(o) \in R$
- The value function v_i is called the **type** of player i
- Designer elicits **types/bids** from players $b = (b_1, \dots, b_n)$
- Designer chooses allocation that maximizes the reported welfare

$$x(b) = \operatorname*{argmax}_{o \in O} RW(o; b) \coloneqq \sum_{i=1}^{n} b_i(o)$$

Total Reported Welfare

Let’s repeat this exercise with two slots

- When highest bidder exists, rest of players achieve reported welfare of …?



General Auction (Mechanism Design) Setting

- Designer chooses allocation that maximizes the reported welfare

$$x(b) = \underset{o \in O}{\operatorname{argmax}}\, RW(o; b) \coloneqq \sum_{i=1}^{n} b_i(o)$$

- Charges to each player their externalities as payment

$$p_i(b) = \underbrace{\max_{o \in O} \sum_{j \neq i} b_j(o)}_{\text{RWelfare of others without me}} - \underbrace{\sum_{j \neq i} b_j\big(x(b)\big)}_{\text{RWelfare of others with me}} \boxed{\geq 0}$$

Why?

How much utility do bidders receive?

- The utility of bidder i for reporting b_i when others report b_{-i}

$$U_i(b) = \underbrace{v_i\big(x(b)\big)}_{\text{My value}} - \underbrace{p(b)}_{\text{My payment}}$$

- If payment=externality

$$U_i(b) = \underbrace{v_i\big(x(b)\big)}_{\text{My value}} - \underbrace{\max_{o \in O} \sum_{j \neq i} b_j(o)}_{\text{RWelfare of others without me}} + \underbrace{\sum_{j \neq i} b_j\big(x(b)\big)}_{\text{RWelfare of others with me}}$$

What is the optimal bid?

- If payment=externality

$$U_i(b) = v_i(x(b)) + \sum_{j \neq i} b_j(x(b)) - \max_{o \in O} \sum_{j \neq i} b_j(o)$$

My value

RWelfare of others with me

RWelfare of others without me

- I want to choose a bid b_i that optimizes my utility

$$\max_{b_i} v_i(x(b)) + \sum_{j \neq i} b_j(x(b)) - \cancel{\max_{o \in O} \sum_{j \neq i} b_j(o)}$$

Does not depend on my bid

What is the optimal bid?

- I want to choose a bid b_i that optimizes my utility

$$\max_{b_i} v_i(x(b)) + \sum_{j \neq i} b_j(x(b))$$

My value RWelfare of others with me

- This only depends on the chosen allocation $x(b)$
- Want to choose a bid that leads to an allocation x that maximizes

$$v_i(x) + \sum_{j \neq i} b_j(x)$$

What is the optimal bid?

- Want to choose a bid that leads to an allocation x that maximizes

$$v_i(x) + \sum_{j \neq i} b_j(x)$$

My value RWelfare of others with me

- Designer chooses allocation that maximizes reported welfare

$$b_i(x) + \sum_{j \neq i} b_j(x)$$

My bid RWelfare of others with me

What is the optimal bid? My true value

- Want to choose a bid that leads to an allocation x that maximizes

$$\underbrace{v_i(x)}_{\text{My value}} + \underbrace{\sum_{j \neq i} b_j(x)}_{\text{RWelfare of others with me}}$$

- Designer chooses allocation that maximizes reported welfare

$$\underbrace{b_i(x)}_{\text{My bid}} + \underbrace{\sum_{j \neq i} b_j(x)}_{\text{RWelfare of others with me}}$$

- If I'm **truthful** then auctioneer chooses the allocation that I want

What is my utility under truthful reporting

- If payment=externality

$$U_i(b) = v_i(x(b)) + \sum_{j\neq i} b_j(x(b)) - \max_{o\in O} \sum_{j\neq i} b_j(o)$$

Total RWelfare with me

RWelfare of others without me

- Since auctioneer optimizes reported welfare:

$$U_i(b) = \max_{o\in O} v_i(o) + \sum_{j\neq i} b_j(o) - \max_{o\in O} \sum_{j\neq i} b_j(o) \boxed{\geq 0}$$

Why?

Total RWelfare with me

RWelfare of others without me

Credits: https://rethinkmedia.org/blog/facebook-instagram-advertising-what-you-need-know-get-started

Learning in Non-Truthful Auctions

Non-Truthful Auctions

- Despite the universality of VCG, non-truthful auctions are frequently used

- More transparent and credible* rules

- The mechanism used in government procurement and display ads

*Credible Auctions: A Trilemma (stanford.edu)

Learning how to bid in auctions

- Given the complexity of digital auction markets
- Given the hardness of strategizing in non-truthful auctions
- Many of these auctions are repeated!

- It makes sense to study learning over time, to decide how to bid

- How do we learn over time when we repeatedly participate in an auction? Can we compete with the best fixed bid in hindsight?

No-Regret Learning in Auctions

At each period $t \in \{1, \dots, T\}$

- An auction among n bidders takes place (GFP, GSP, FP)
- Each bidder i submits bid b_i from discrete set of N bids $\{\epsilon, 2\epsilon \dots, 1\}$
- Each bidder learns their allocation and payment

$$x_i^t, p_i^t = x_i(b^t), p_i(b^t)$$

- e.g. in a first price auction, learn whether I won
- e.g. in a second price auction, learn whether I won and when I win, I learn the next highest bid.

No-Regret Learning

- Want to choose my bids b_i^t, based on algorithm that guarantees

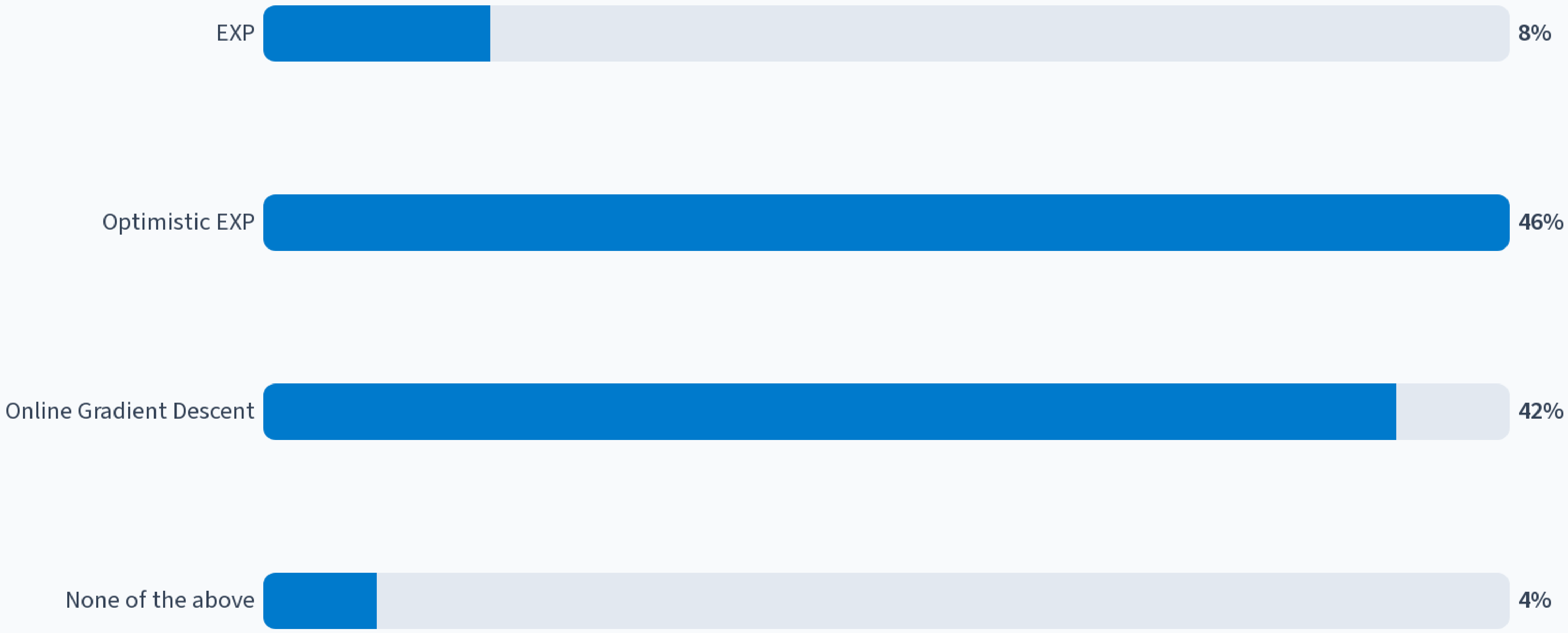
$$\frac{1}{T}\sum_{t=1}^{T} u_i\left(b^t\right) \geq \max_{b_i \in [N]} \frac{1}{T}\sum_{t=1}^{T} u_i\left(b_i, b^t\right) - \epsilon(T)$$

- for some $\epsilon(T) \to 0$

What algorithm should I use?

- EXP
- Optimistic EXP
- Online Gradient Descent
- None of the above

What algorithm should I use?



What algorithm should I use?

No-Regret Learning with Limited Feedback

- Want to choose my bids b_i^t, based on algorithm that guarantees

$$\frac{1}{T}\sum_{t=1}^{T} u_i(b^t) \geq \max_{b_i \in [N]} \frac{1}{T}\sum_{t=1}^{T} u_i(b_i, b^t) - \epsilon(T)$$

- Seems like a standard N action no-regret problem

- What's the catch! I don't receive after each period the utility for all my actions. Only the utility for action I took!
- Limited Feedback. I cannot calculate how much I would have gotten with any other bid (e.g. in an FP, solely knowing whether I won or not).

No-Regret Learning with Bandit Feedback

At each period t

- Adversary chooses a loss vector $\ell_t \in [0,1]^N$
- I choose an action i_t (not knowing ℓ_t)
- I observe loss of my chosen action $\ell_t^{i_t}$

- I want to guarantee small expected regret with any fixed action:

$$\max_{i \in N} E\left[\frac{1}{T}\sum_{t=1}^{T} \ell_t^{i_t} - \ell_t^{i}\right] \leq \epsilon(T)$$

Constructing Un-biased Estimates of Vector

- There is a hidden loss vector $\ell_t = \left(\ell_t^1, \dots, \ell_t^N\right)$ (potential outcomes)
- At each period I choose action (treatment) j with probability p_t^j
- I learn the loss ℓ_t^j with probability p_t^j
- Remember: no-regret algorithms work well, even if we have unbiased proxies of the true losses (e.g. Monte Carlo CFR)

Question. Can I construct a random variable that guarantees that in expectation over the choice of actions?

$$E\left[\tilde{\ell}_t\right] = \ell_t \Leftrightarrow \forall j\colon\ E\left[\tilde{\ell}_t^j\right] = \ell_t^j$$

Constructing Un-biased Estimates of Vector

Question. Can I construct a random variable that guarantees that in expectation over the choice of actions?

$$E\left[\tilde{\ell}_t\right] = \ell_t \Leftrightarrow \forall j:\ E\left[\tilde{\ell}_t^j\right] = \ell_t^j$$

- Random variable can always depend on identity of chosen action j_t. When I choose j random variable can also depend on ℓ_t^j

$$\tilde{\ell}_t^j = 1\{j_t = j\} f_j\left(\ell_t^j\right) + 1\{j_t \neq j\} g_j(j_t)$$

- Let's make g_j zero, and f_j linear in ℓ_t^j

$$\tilde{\ell}_t^j = 1\{j_t = j\} a_j \ell_t^j \Rightarrow E\left[\tilde{\ell}_t^j\right] = p_t^j a_j \ell_t^j = \ell_t^j \Rightarrow a_j = \frac{1}{p_t^j}$$

Inverse Propensity Estimates

At each period t

- Consider the random variables

$$\tilde{\ell}_t^j = \frac{1\{j_t = j\}}{p_t^j} \ell_t^j$$

- The vector $\tilde{\ell}_t$ can always be calculated $\left(0, \ldots, 0, \frac{\ell_t^{j_t}}{p_t^{j_t}}, 0, \ldots, 0\right)$
- The vector $\tilde{\ell}_t$ is an unbiased proxy of the true loss vector:

$$E\left[\tilde{\ell}_t\right] = \ell_t$$

The EXP Algorithm with Bandit Feedback

```
Initialize pt to the uniform distribution
For t in 1..T
    Draw action jt based on distribution pt
    Observe loss of chosen action lt[jt]

    Construct un-biased proxy loss vector
      ltproxy[j] = 1(jt=j) * lt[jt] / pt[jt]

    Update probabilities based on EXP update
      pt = pt * exp(-eta * ltproxy)
      pt = pt / sum(pt)
```

Recap: Regret of FTRL

(FTRL) $$x_t = \operatorname*{argmin}_{x \in X} \boxed{\sum_{\tau < t} \langle x, \ell_\tau \rangle} + \boxed{\frac{1}{\eta} \mathcal{R}(x)}$$

1-strongly convex function of x that stabilizes the maximizer

Historical performance of always choosing strategy x

Theorem. Assuming the utility function at each period

$$f_t(x) = \langle x, \ell_t \rangle$$

is L-Lipschitz with respect to some norm $\|\cdot\|$ and the regularizer is 1-strongly convex with respect to the same norm then

$$\text{Regret} - \text{FTRL}(T) \le \boxed{\eta L} + \boxed{\frac{1}{\eta T}\left(\max_{x \in X} \mathcal{R}(x) - \min_{x \in X} \mathcal{R}(x)\right)}$$

Average stability induced by regularizer

Average loss distortion caused by regularizer

Problem! The loss vector $\tilde{\ell}_t$ is not in [0,1].

It can take huge values, as probability of an action goes to 0!

Intuition: if probability goes to 0, then this action is chosen very infrequently. The loss vector very rarely takes this large value, i.e., the *variance* of the loss should be small.

Variance of Loss Vector

- Variance is

$$E\left[\left(\tilde{\ell}_t^j\right)^2\right] - E\left[\tilde{\ell}_t^j\right]^2 = E\left[\left(\tilde{\ell}_t^j\right)^2\right] - E\left[\ell_t^j\right]^2$$

- Second term is in $[0, 1]$. We will focus on first term (call it "variance")

$$E\left[\left(\tilde{\ell}_t^j\right)^2\right] = p_t^j \left(\frac{\ell_t^j}{p_t^j}\right)^2 = \frac{\left(\ell_t^j\right)^2}{p_t^j}$$

- And we collect this "variance" term only when end up choosing j

$$\text{Average "Variance"} = \sum_j p_t^j \cdot E\left[\left(\tilde{\ell}_t^j\right)^2\right] = \sum_j \left(\ell_t^j\right)^2 \leq N$$

Recap: Regret of FTRL

(FTRL) $$x_t = \operatorname*{argmin}_{x \in X} \boxed{\sum_{\tau<t} \langle x, \ell_\tau \rangle} + \boxed{\frac{1}{\eta} \mathcal{R}(x)}$$

1-strongly convex function of x that stabilizes the maximizer

Historical performance of always choosing strategy x

Can we replace L with the Average "Variance"?

Theorem. Assuming the utility function at each period

$$f_t(x) = \langle x, \ell_t \rangle$$

~~is L-Lipschitz with respect to some norm $\|\cdot\|$~~ and the regularizer is 1-strongly convex with respect to the same norm then

$$\text{Regret} - \text{FTRL}(T) \le \boxed{\eta \boldsymbol{L}} + \boxed{\frac{1}{\eta T} \left(\max_{x \in X} \mathcal{R}(x) - \min_{x \in X} \mathcal{R}(x) \right)}$$

Average stability induced by regularizer

Average loss distortion caused by regularizer

Update: Regret of EXP

(EXP) $$p_t = \operatorname*{argmin}_{p \in \Delta} \sum_{\tau < t} \langle p, \tilde{\ell}_\tau \rangle + \boxed{\frac{1}{\eta} \mathcal{R}(p)} \begin{pmatrix} \text{Negative} \\ \text{Entropy} \end{pmatrix} \quad \mathcal{R}(p) = \sum_{i=1}^{n} p_i \log(p_i)$$

$$p_t \propto p_{t-1} \exp\left(-\eta\, \tilde{\ell}_{t-1}\right)$$

Theorem. Assuming $\tilde{\ell}_t$ are random proxies that, conditional on history, have expected value equal to true loss vector ℓ_t and $\tilde{\ell}_t \geq 0$, then regret of EXP is bounded as:

$$\text{Regret} - \text{EXP}(T) \leq \frac{\eta}{T} \sum_t E\left[\sum_j p_t^j \left(\tilde{\ell}_t^j\right)^2\right] + \frac{\log(N)}{\eta T}$$

See Lemma 20 and Theorem 2.22 and Corollary 4.2 of mal-018.dvi (huji.ac.il)

Update: Regret of EXP

(EXP) $$p_t = \operatorname*{argmin}_{p\in\Delta} \sum_{\tau<t} \langle p, \tilde{\ell}_\tau \rangle + \boxed{\frac{1}{\eta}\mathcal{R}(p)} \begin{pmatrix}\text{Negative}\\ \text{Entropy}\end{pmatrix} \quad \mathcal{R}(p) = \sum_{i=1}^{n} p_i \log(p_i)$$

$$p_t \propto p_{t-1} \exp\left(-\eta\, \tilde{\ell}_{t-1}\right)$$

Theorem. Assuming $\tilde{\ell}_t$ are random proxies that, conditional on history, have expected value equal to true loss vector ℓ_t and $\tilde{\ell}_t \geq 0$, then regret of EXP is bounded as:

$$\text{Regret} - \text{EXP}(T) \leq \frac{\eta}{T} \sum_t E\left[\sum_j p_t^j\, E\left[\left(\tilde{\ell}_t^j\right)^2\right]\right] + \frac{\log(N)}{\eta T}$$

Expected Average
“Variance”?

Update: Regret of EXP

(EXP) $$p_t = \operatorname*{argmin}_{p \in \Delta} \sum_{\tau < t} \langle p, \tilde{\ell}_\tau \rangle + \boxed{\frac{1}{\eta} \mathcal{R}(p)} \left(\begin{array}{c} \text{Negative} \\ \text{Entropy} \end{array} \right) \mathcal{R}(p) = \sum_{i=1}^{n} p_i \log(p_i)$$

$$p_t \propto p_{t-1} \exp\left(-\eta\, \tilde{\ell}_{t-1}\right)$$

Theorem. Assuming $\tilde{\ell}_t$ are random proxies that, conditional on history, have expected value equal to true loss vector ℓ_t and $\tilde{\ell}_t \geq 0$, then regret of EXP is bounded as:

$$\text{Regret} - \text{EXP}(T) \leq \frac{\eta}{T} \sum_t N + \frac{\log(N)}{\eta T}$$

For the inverse propensity proxies

Update: Regret of EXP

(EXP) $$p_t = \underset{p \in \Delta}{\operatorname{argmin}} \sum_{\tau < t} \langle p, \tilde{\ell}_\tau \rangle + \boxed{\frac{1}{\eta} \mathcal{R}(p)} \begin{pmatrix} \text{Negative} \\ \text{Entropy} \end{pmatrix} \mathcal{R}(p) = \sum_{i=1}^{n} p_i \log(p_i)$$

$$p_t \propto p_{t-1} \exp\left(-\eta\, \tilde{\ell}_{t-1}\right)$$

Theorem. Assuming $\tilde{\ell}_t$ are random proxies that, conditional on history, have expected value equal to true loss vector ℓ_t and $\tilde{\ell}_t \geq 0$, then regret of EXP is bounded as:

$$\text{Regret} - \text{EXP}(T) \leq \eta N + \frac{\log(N)}{\eta T}$$

For the inverse propensity proxies

Update: Regret of EXP

$$\text{(EXP)} \qquad p_t = \underset{p\in\Delta}{\operatorname{argmin}} \sum_{\tau<t} \langle p, \tilde{\ell}_\tau \rangle + \boxed{\frac{1}{\eta}\mathcal{R}(p)} \quad \left(\begin{array}{c}\text{Negative}\\ \text{Entropy}\end{array}\right) \quad \mathcal{R}(p) = \sum_{i=1}^{n} p_i \log(p_i)$$

$$p_t \propto p_{t-1} \exp\left(-\eta\, \tilde{\ell}_{t-1}\right)$$

Theorem. Assuming $\tilde{\ell}_t$ are random proxies that, conditional on history, have expected value equal to true loss vector ℓ_t and $\tilde{\ell}_t \geq 0$, then regret of EXP is bounded as:

$$\text{Regret} - \text{EXP}(T) \leq \eta N + \frac{\log(N)}{\eta T} \Rightarrow \text{Regret} - \text{EXP}(T) \lesssim \sqrt{\frac{N\log(N)}{T}}$$

For $\eta \sim \sqrt{\frac{\log(N)}{NT}}$

See Lemma 20 and Theorem 2.22 and Corollary 4.2 of mal-018.dvi (huji.ac.il)