Cheating in Ranking Systems

Paper: Lihi Dery, Dror Hermel, Artyom Jelnov 2020

Presenter: Will Ottenheimer

Contents

Background

Consider a marketplace where products are ranked

Rankings are powerful enough to create the desire to cheat

Marketplace management has incentive to police ranking system

Key question:

“How do detection quality and product rankings affect incentive to cheat?”

Background - continued

Two players - Application (App) & Platform

High ranking leads to short term profit - App

Marketplace needs to manage its reputation - Platform

Related Work

Inspection games Avenhaus (2002) - Key difference is automatic violation signal rather than costly detection decision

Economics of law enforcement Becker (1968) - This paper is different because rank manipulation is a specific kind of violation where enforcer and potential violator may share interests

Tort Law Landes and Posner (1984) - different because platform is not compensated for damages.

Motivation

Assumption: Detection is possible

Existing literature says algorithms can pick up patterns in fake reviews and big shifts in ratings

Assumption: Cheating can work

Manipulation firms exist and can generate drastic ranking changes

“crowdturfing”

Motivation

Assumption: Rank => Profit

Existing literature says that rankings matter

You need ranking for discovery also rank increases willingness to pay

Platform marketplace is structurally different than supermarket

Model - 2 players & 2 settings

Application - Potential to violate ranking system

Platform - charges commission fee and monitors imperfectly (at no cost)

Exogenous fee

Endogenous fee

Special case for the topmost rating

Model - Preliminaries

Risk neutral agents

Ratings $r \in [0,1]$ Uniform distribution

Notation	Description
A	Application
P	Platform
r	A's rating
f	Proportional fee that is paid by the application to the platform
γ	Coefficient that relates A's rating to A's profit
c	Action of cheating
$\hat{c}$	Action of not cheating
e	Cost of cheating
d	Lower limit of the rating, following cheating
s	Suspecting alert
$\hat{s}$	Non-suspecting alert
b	Banning action
$\hat{b}$	Non-banning action

α	Probability of a type-I error in detection
β	Probability of type-II error in detection
v	Revenue from not making false accusations
w	Cost of non-detection
P_c	Probability of cheating
$\bar{r}$	Threshold of suspicious rating, determined by the platform
ρ	Initial rating in the topmost model
$\alpha(\rho)$	Probability of a type-I error in detection in the topmost model
$\beta(\rho)$	Probability of type-II error in detection in the topmost model
$l(\rho)$	The probability to obtain honestly the highest rating (in the topmost model)
P_b	Probability that P bans suspected A (in the topmost model)

Model - Payoffs

Application revenue:

No Cheat: $\gamma r(1 - f)$ Cheat: $\gamma r(1 - f) - e$

Platform utility:

Application cheats and is not banned: $\gamma rf - w$

No false accusation: $\gamma rf + v$

Model - errors

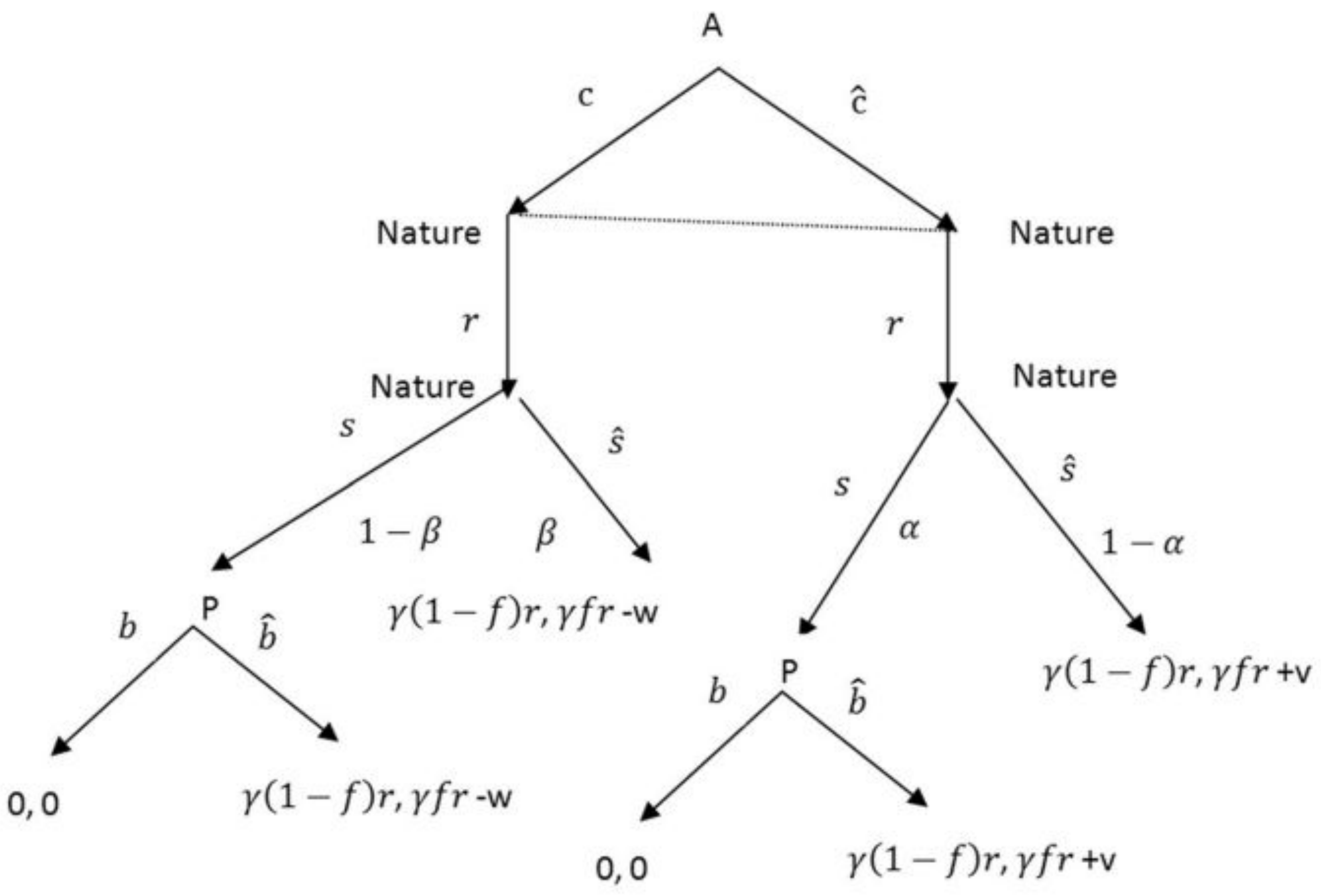
Type - I

Platform detects cheating when there is no cheating

Type - II

Platform doesn’t detect cheating when application cheats

*Error probability given at start based



A
c
ĉ
Nature
Nature
r
r
Nature
Nature
s
ŝ
s
ŝ
1 − β
β
α
1 − α
P
b
b̂
γ(1 − f)r, γfr -w
γ(1 − f)r, γfr +v
P
b
b̂
0, 0
γ(1 − f)r, γfr -w
0, 0
γ(1 − f)r, γfr +v

Model - Exogenous Fee

Proposition 1

$\alpha > 0$ Type I errors are possible

1) If $w < \gamma f d$ Then application cheats and platform doesn’t ban

   Explanation

   Not banning is dominant since fee covers reputation loss

Model - Exogenous Fee

Proposition

2) If $w > \gamma fd$ and $\beta > [d+(1-\alpha)(1+d)] / (1+d)$ then App always cheats and platform always bans

Explanation

When Type II error is high cheating is more rewarding.

Reputation effect outweighs fee effect

Model - Exogenous Fee

Proposition 1

3) If w is high enough and $\beta < [(1-\alpha)(1-d^2)+d^2] / (1+d)$ $\exists$ equilibrium where app cheats with probability Pc and after violation signal platform bans iff

$r > \bar{r}, \; d < \bar{r} < 1$

Produces corollary 1

Model - Exogenous Fee

Corollary 1

1. Probability app cheats increases in f and v… decreases in w
2. Threshold r decreases in α and in ß

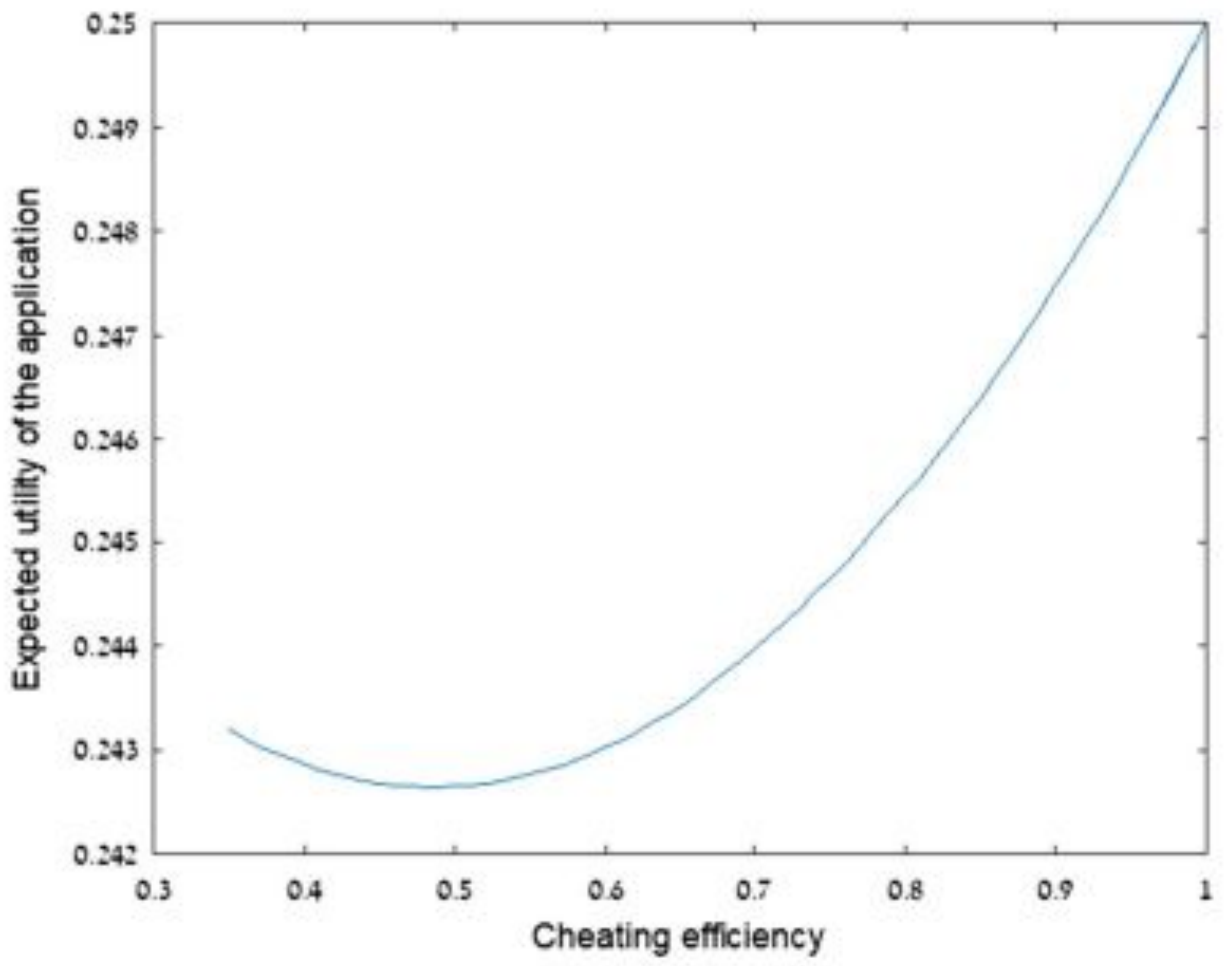
When prob of false accusation rises app has more incentive to cheat (get caught anyway)

When more incentive to cheat platform widens ban region

Alternatively probability of non detection (increased ß) encourages cheating and ban region increases

Efficiency of Cheating

As minimum rating after cheating rises payoffs are higher, however this also makes platform more suspicious



Model - Endogenous Fee

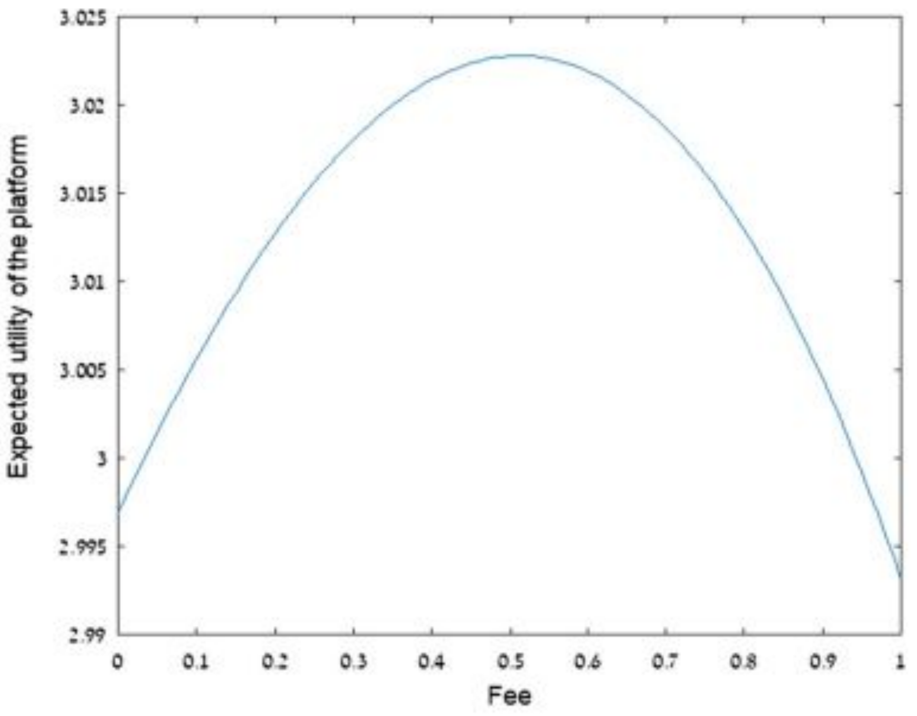
Suppose initial stage now firm sets fee.

Higher F increases platform profits but also encourages cheating

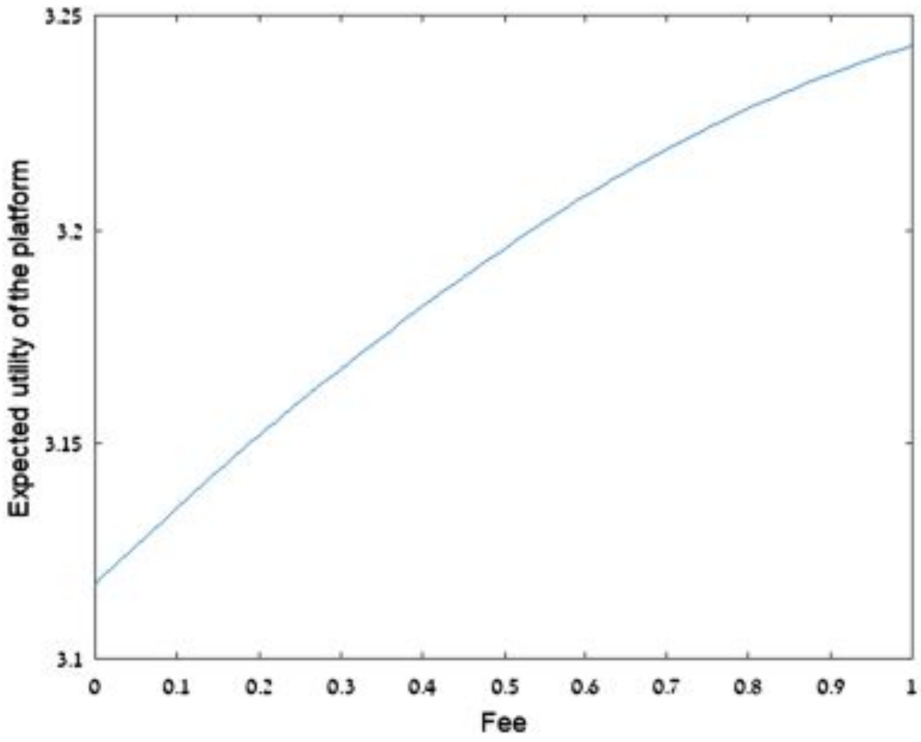
Platform utility maximizing fee is dependent on reputation w

Model- Endogenous Fee

W = 3



W = 4



Model - Topmost rating

Now cheating guarantees top rating 1 and detection is dependent on initial rating of topmost model

Model - Topmost rating

Proposition 2

1) If $w < \gamma f d$ Then application cheats and platform doesn’t ban

   Explanation

   Not banning is dominant since fee covers reputation loss

Model - Topmost rating

Proposition 2

2) $l(p) - \alpha(p)l(p) + p - pl(p) < \beta(p)$ then application cheats and platform bans after alert s

Explanation: This is condition where payoff of pure cheat dominates not cheat

Model - Topmost rating

Proposition 2

3) If $w > \gamma fd$ and $l(p) - \alpha(p)l(p) + p - pl(p) > ß(p)$ then application cheats with probability Pc and platform bans with probability Pb

Produces Corollary 2

Model - Topmost rating

Corollary 2

3) let $w > \gamma f d$ and $l(p) - \alpha(p) l(p) + p - p l(p) > \beta(p)$

Then probability of cheating increases in f, v, $\alpha(p)$ and $\beta(p)$... decreases in w

The probability Platform bans is increasing in $\alpha(p)$ and $\beta(p)$

Unreliable detection => lots of cheating => lots of banning

Model - Topmost rating

Corollary 3

3) let $\alpha(p)$ be constant = α and $w > f\alpha$ and $l(p) - \alpha(p)l(p) + p - pl(p) > \beta(p)$

Then cheating increases in initial ranking p

When ranking is closer to 1 initially and suspicions of top ranking is lower, then apps cheat more freely

Conclusions

Significant findings: Higher fee leads to more cheating

- A monopolistic platform might not desire a higher fee

Better detection (lower error probabilities) lowers cheating

- Suggests platforms advertise detection capabilities

Results are based on cost of cheating being negligible and exogenous reputation