The Incentive to Report Sexual Harassment: A Game Theoretic Approach

by

Jingqing Liu

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Bachelor of Arts, Honours
in the Department of Economics
University of Victoria
April 2018

Supervised by Dr. Chris Auld and Dr. Chris Willmore

For

Dr. Elisabeth Gugl, Honours co-advisor

Dr. Martin Farnham, Honours co-advisor

Abstract

This paper studies the incentive to truthfully report sexual harassment, as well as the incentive to make a false report in a two-stage game theoretic model. By assuming that victims and liars face the similar payoff structure, I find that victims and liars' equilibrium strategy is very similar. In lack of physical evidences, it is hard for the public (authority) to completely separate true reports from false ones. An online reporting system Callisto that allows victims to file up an electronical report and only forward it to the authority if and only if there exists another report against the same harasser also gives liars a greater incentive to make a false report. If the authority's decision is made only depending the number of report in the system, the existence of Callisto reporting system reduce the probability of letting go a harasser who commit a crime (type II error) but increase the probability of punishing the innocent (type I error).

Key Words: Sexual Harassment Reporting, Information Escrow, Project "Callisto", False Report

Table of Contents

1. Introduction

This paper studies the incentive of reporting sexual harassment using economic theory. Sexual harassment refers to the "behavior characterized by the making of unwelcome and inappropriate sexual remarks or physical advances in a workplace or other professional or social situation." In most modern legal context, sexual harassment is illegal (Oxford Living Dictionary, 2018). However, the lack of physical evidence lead to many difficulties in uncovering the truth of sexual harassment.

The authority (or public)'s action is heavily depending on the report of sexual harassment. The reporting of sexual harassment has been deeply studied by psychologists, sociologists and legal scholars but rarely by economists. Since incentives likely play an important role in reporting sexual harassments, it makes sense to consider the problem from an economic perspective.

In this paper, I focus on the following three problems of sexual harassment reporting that are mostly mentioned and documented in the media and literatures:

1. *Not every victim is willing to come forward and make a report.*
2. *Some reports are significantly delayed.*
3. *Misreporting (error) occurs*.

Many studies have shown that sexual harassment is highly underreported in North America (Benoit et al 2015, Brennan and Taylor-Butts 2008, Kaufman 2008, Luce et al 2010). Moreover, given the number of report we have, a great proportion of them relate to incidences that happened many years ago. The statistics from The Rape, Abuse & Incest National Network (RAINN) shows that "among those victims who do report, they wait an average of 11 months."

A real-world example is Harvey Weinstein's sexual harassment case in Hollywood. As a former film producer and the co-founder of Weinstein company, Harvey Weinstein was accused of more than 12 incidents of sexual harassment allegations over a at least 30-year-period (Wikipedia, 2017). After The New York Times and The New Yorker reported these accusations to the public in October 2017, more than 80 women in the film industry came forward and made similar accusations (Victor, 2017). In 2014, *Rolling Stone* published an article, "A Rape on Campus," accusing three Duke Lacrosse team players of gang rape two years before. However, in 2015 the magazine had to retract the article because it was discovered that the rape was fabricated (Fowler, 2014).

A small number of economics studies in this area (Ayres and Unkovic 2012, Lee and Sueng 2017) focus on the problems I addressed in this paper. Ayres and Unkovic(2012) primarily study the underreporting issue, and suggests that due, in part, to underreporting by others, a sexual harassment victim often doesn't know the identity of other victims, or whether other victims even exist. In such circumstances, it is often the case that "no one moves unless everyone moves, so no one moves." (Eisenmann, 2006). Sometimes, a habitual sexual harasser may never be reported because no one is willing to make the first (and possibly the only) report (Ayres and Unkovic). In addition, Ayres and Unkovic suggests that the idea of information escrow could be used to encounter the underreporting problem in sexual harassment reporting. Information escrow usually means "an entity that hold on information for you and only release it to a third party when certain pre-agreed upon conditions are met."

As a type of information escrow (Ayres and Unkovic, 2012), a website based reporting system (Callisto) which allows victims "to choose to have their evidence forwarded to authorities automatically once another person uploads evidence about the same assailant"

is built to incentivize more reports and combat sexual harassment. According to the website's founder and CEO, Jessica Ladd, Callisto is a website based reporting system that creates opportunity for sexual assault or harassment victims to "electronically report their harassment with an option to create a secure, timestamped document of what happened to them… preserving the evidence even they don't want to report yet, and most critically, with the ability to report their harassment, only if someone else reports the same assailant."

Building on Ayres's (2012) research, Lee and Sueng (2017) developed a two-stage dynamic model where victims and libelers strategically choose whether and when to allege a crime. In Lee and Sueng's model, the potential harasser is either a "good" or a "bad" type. The bad type will potentially harass a person with a given probability, while there is another probability that a player will form a grudge against good type ad then falsely claim that they have been harassed. If the harasser is a bad type but he doesn't harass, then nothing happens. The authors denote the person harassed by the harasser as victim, and denote the person who forms the grudge as libelers. Victim and libeler face the exactly same payoff and cost structure. The probability of a bad type harasses a player is larger than the probability of a player forms grudge in each stage. As a result, Lee and Sueng (2017) suggested that without any hard evidence, the public cannot completely separate true victims from libelers by the strategy they choose in the equilibrium. They prove that of "waiting for corroboration" (delaying report) does emerge as an equilibrium behavior. The public usually assigned less credibility to an allegation against a distant incidence than a recent one, and a single report is usually not convincing enough for the public (authority) to make sanction decision.

In this paper, I modified the basic setup of Lee and Sueng (2017)'s model and change assumptions based on their results. Firstly, instead of assigning two types to the potential

harasser and only a bad type could potentially harass a player, I simple assign a probability that harasser will harass regardless of what type he is. In another word, instead of saying that (for example) there are 50% people in our society is good and 50% people is bad and only those bad ones will harass with a certain probability, I assume that we live in a world where everyone has a probability to harass another person. Secondly, Lee and Sueng only allow player to form grudge and falsely report when the harasser is a good type. I assigned a probability that person may be motivated by some financial gain and make a false report when the harasser does has a probability to harass but he doesn't harass.

By changing the basic set up of the model, this paper compares the results with conclusion Lee and Sueng reached. More specifically, this paper answers following 2 research question: 1) Given the lack of hard evidence, can the public (authorities) discriminate false allegation from the ones that report a true incidence? (i.e. can the public or authority separate true victims from "liars" by the strategy they choose in the equilibrium? 2) How the existence of Callisto reporting system changes the type I error and reduces type II error comparing to the equilibrium outside the Callisto system.

2. Model Setup

There is an agent ***A*** (the potential harasser) who is active for two periods. ***A*** enters both stages and interacts with player 1 in stage 1 and player 2 in stage 2. Player 1 enters both stages while player 2 only enters stage 2.

In each period, ***A*** harass a player with independent probability *p*. If a player is harassed by ***A***, then this player is a victim. If a player is not harassed by ***A***, then there is a probability of *q* that this player form a grudge against ***A*** (or be motivated by some financial gain to make a false report) and becomes a liar

The state of the world is whether harassment takes place. Both victim and liars gains utility from getting $\boldsymbol{A}$ arrested by the authority. The payoff is 1 if $\boldsymbol{A}$ is arrested and 0 if $\boldsymbol{A}$ is not arrested. Given reporting a sexual harassment is costly, I denote the cost of reporting by c_v for victim and c_l for liars, and assume both c_v and c_l randomly takes value from the range [0,1] that is uniformly distributed. The cost of reporting is a private information that is only know by the reporter.

For $t = 1,2$, we let V_t represent the victim harassed by $\boldsymbol{A}$ in period *t* and L_t represent the liars who form grudge (or incentivised by some financial gain) against $\boldsymbol{A}$ in period *t.* Without previous report, a player does not know whether she lives in stage 1 or stage 2. As Lee and Sueng (2017) did in their paper, I assign an ***ex ante*** probability of 0.5 two each of these two possibilities.

Player 1 in stage 1 has two chances to report the harassment. Player 1 in stage 1 can either report immediately or make a delayed report in stage 2 followed by player 2's report in stage 2. V_1 is harassed by $\boldsymbol{A}$ in stage 1, she can report the harassment immediately in stage 1 with probability $\alpha(c_v)$. Second, if V_1 does not report in period one, she can make a delayed report at the end of period 2, after observing whether another victim has made a report against $\boldsymbol{A}$. Let $\hat{\alpha}_1(c_v)$ represent the probability that that V_1 make a delayed report in period 2. A victim applies a discount factor of 0.5 to benefits that are obtained or costs that are incurred a stage later. L_1 is motivated to make a false report against $\boldsymbol{A}$ in stage 1. L_1 reports immediately with probability $\beta(c_l)$ and makes a delayed report with a probability $\hat{\beta}_1(c_l)$. A liar doesn't apply any discount factor to the benefits nor costs a stage later.

Player 2 in stage 2 only has one chance to make a report. V_2 is harassed by $\boldsymbol{A}$ in period 2. If no one has accused $\boldsymbol{A}$ before, V_2 does not know which stage she lives in. In this case, her

strategy is the same as V_1. She can report the harassment immediately in period one with probability $\alpha(c_v)$. If some has reported, then V_2 observed this report and make another report worth probability of after observing whether another victim has made a report against ***A***. Let $\hat{\alpha}_2(c_v)$ represent the probability that that V_2 reports if there is a prior report in stage 1. L_2 is motivated to make a false report against ***A*** in stage 2. L_2 reports immediately with probability $\beta(c_l)$ and makes a delayed report with a probability $\hat{\beta}_2(c_l)$ if there is a report previously. I define $a = E[\alpha(c_v)]$, and $b = E[\beta(c_l)]$, where the expectation is taken over possible realization of $c_v(c_l)$.

The authority decides whether to arrest ***A*** after the following observable outcome: (1) there was only one report in stage 1. (I denote this event by *rn*); (2) there was only one report in stage 2 (I denote this event by *nr*); (3) there was no report. (I denote this event by *nn*); (4) there was no repot in stage one but two report in stage 2 (which represent that person 1 delayed the report) (I denote this event by *rR*). (5) there are one report in stage 1 and another report in stage 2. (I denote this event by *rr*). Let *w* denotes the actual outcome that authority observe. For $w \in \{rn, nr, nn, rR, rr\}$, let $\hat{p}(w)$ represent the probability that authority think ***A*** independently harass a person. A will be arrested of $\hat{p}(w) \geq s \in (0,1)$, where *s* is the exogenous public standard threshold.

Assumption 1. *Potential harasser **A**'s behavior is non-strategic.*

This assumption is made to 1) simplify the model and 2) keep the focus at the incentive to report. Therefore, the real strategic players in this game is player 1 in stage 1 and player 2 in stage 2.

Assumption 2. $\hat{p}(rr, rR) \geq s$*; and moreover* $\hat{p}(nn, rn, nr) < s$

*Assumption 2 is saying that only two report (either event rr to rR) lead to arrest. One single report from either period is not enough to arrest **A**. Since the event nn doesn't give the authority any information, **A** will not be arrested given no reports.*

3. Equilibrium Strategy

3.1 Equilibrium Strategy Outside Callisto System

Firstly, consider a person who has observed a pervious report. Given **Assumption 2.**, a silent victim V_1 who observes another person making a report against ***A*** in stage 2 will surely lodge a delayed report, because doing so gets ***A*** arrested and gives her a payoff of 1 at a cost $c_v < 1$. Similarly, a victim V_2 who knows of a previous report against A will also make a report. Since the payoff structure of liars is identical to the payoff structure of victim, we know that a silent liar L_1 who observes another person making a report against a in stage 2 will surely make a delayed report. A liar L_2 who knows of a previous report against A will also make a report.

Proposition 1. *Person 2 in stage 2 always reports if there is a report in stage 1. Person 1 will always make a delayed report at the end of stage 2 if person 2 report immediately. The equilibrium strategy is therefore:* $\hat{\alpha}_1(c_v) = \hat{\alpha}_2(c_v) = 1$ *for any* c_v*;* $\hat{\beta}_1(c_l) = \hat{\beta}_2(c_l) = 1$ *for any* c_l.

Secondly, consider now a person who has not seen a prior accusation against ***A.*** For convenience, we called this person a *fresh victim* or a *fresh liar* depending on their type. A fresh victim who has not seen a prior report against ***A*** can entertain following 6 events with corresponding probabilities:

1. She is V_1, there is another victim exist in the next stage. (with probability ***0.5p***)

2. She is V_1, there is a liar exist in the next stage. (with probability ***0.5(1-p)q***)
3. She is V_1, neither liars or victim exist in the next stage.

   (with probability ***0.5(1-p)(1-q)***)
4. She is V_2, there is a victim exist in stage 1 but she didn't report.

   (with probability ***0.5p(1-a)***)
5. She is V_2, there is a liar exist in stage 1 but she didn't report.

   (with probability ***0.5(1-p)q(1-b)***)
6. She is V_2, neither liars or victim exist in the next stage.

   (with probability ***0.5(1-p)(1-q)***)

The sum of above probability is: *1-0.5ap-0.5bq+0.5pqb*

Therefore, if a fresh victim has a reporting cost of c_v, the expected net payoff from reporting immediately is:

$$-c_v + \frac{0.5p + 0.5(1-p)q}{1 - 0.5ap - 0.5bq + 0.5pqb} 0.5 + \frac{0.5p(1-a) + 0.5(1-p)q(1-b)}{1 - 0.5ap - 0.5bq + 0.5pqb}$$

The expected net payoff from not report immediately is:

$$0.5(1-c_v)\frac{0.5pa + 0.5(1-p)q}{1 - 0.5ap - 0.5bq + 0.5pqb}$$

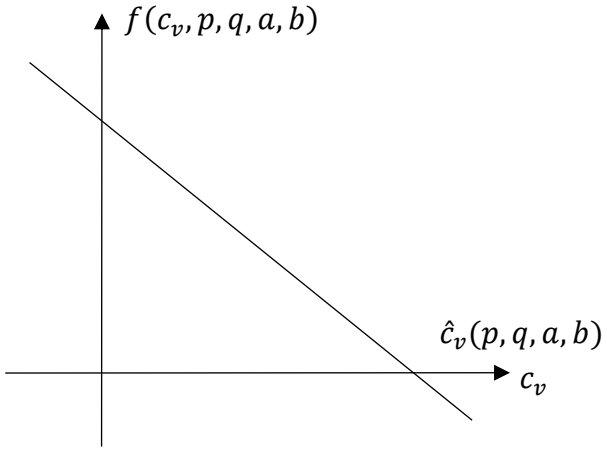
Let $f(c_v, p, q, a, b)$ represent the payoff difference between reporting immediately and not report immediately for a fresh victim, given that other person adopting the strategy in **Proposition 1.**.

Equilibrium requires the net expected payoff from report immediately equals to the net expected payoff from not reporting immediately (i.e. $f(c_v, p, q, a, b) = 0$.) For any $p, q, a, b,\ f(c_v, p, q, a, b)$ is a strictly decreasing function in c_v, because:

$$\frac{\partial f(c_v, p, q, a, b)}{\partial c_v} < 0$$

Figure1. below provides a graphical illustration of the relationship between the difference function and the report cost of a fresh victim. For any $c_v < \hat{c}_v(p, q, a, b)$, $f(c_v, p, q, a, b) > 0$. The net expected payoff of reporting immediately is larger than the net expected payoff of not doing so. For any $c_v > \hat{c}_v(p, q, a, b)$, $f(c_v, p, q, a, b) < 0$. The net expected payoff of reporting immediately is smaller than the net expected payoff of not doing so. If $c_v = \hat{c}_v(p, q, a, b)$, then the payer will be indifference between reporting immediately and not reporting immediately.

Figure 1.



Proposition 2. There exists a unique threshold cost $\hat{c}_v(p, q, a, b)$ satisfying $f(\hat{c}_v(p, q, a, b), p, q, a, b) = 0$. The best response for fresh victim is to report immediately (i.e. choose $\alpha(c_v)$=1) if and only if $c_v \leq \hat{c}_v(p, q, a, b)$.

In the similar fashion, if a fresh liar who has not seemed a prior report against A, she then can entertain those 6 condition with a corresponding probability just like the fresh victim:

1. She is L_1, there is another victim exist in the next stage. (with probability ***0.5p***)

2. She is L_1, there is a liar exist in the next stage. (with probability ***0.5(1-p)q***)
3. She is L, neither liars or victim exist in the next stage.

   (with probability ***0.5(1-p)(1-q)***)
4. She is L_2, there is a victim exist in stage 1 but she didn't report.

   (with probability ***0.5p(1-a)***)
5. She is L, there is a liar exist in stage 1 but she didn't report.

   (with probability ***0.5(1-p)q(1-b)***)
6. She is L_2, neither liars or victim exist in the next stage.

   (with probability ***0.5(1-p)(1-q)***)

the expected net payoff from reporting immediately is:

$$-c_l + \frac{0.5p + 0.5(1-p)q + 0.5p(1-a) + 0.5(1-p)q(1-b)}{1 - 0.5ap - 0.5bq + 0.5pqb}$$

the expected net payoff from not report immediately is:

$$(1-c_l)\frac{0.5pa + 0.5(1-p)q}{1 - 0.5ap - 0.5bq + 0.5pqb}$$

Let $f(c_l, p, q, a, b)$ represent the payoff difference between reporting immediately and not report immediately for a fresh liar, given that other person adopting the strategy in **Lemma 1.**. The exact same procedure is procced as I did for the fresh victim.

Proposition 3. *There exists a unique* $\hat{c}_l(p, q, a, b)$ *satisfying* $f(\hat{c}_l(p, q, a, b), p, q, a, b) = 0$, *such that the best response for fresh liar is to report immediately (i.e. choose* $\alpha(c_l)$*=1) if and only if* $c_l \leq \hat{c}_l(p, q, a, b)$.

3.2 Equilibrium Strategy Under Callisto Reporting System

In this section, I examine the outcome assuming that authority uses Callisto as the only reporting channel, while keeping the basic setup of my model the same. I assume that an no cost is incurred when one enters a report into Callisto system and a cost *c* the report is submitted to the authority.

Since the system will show no report to the authority anytime when there is one report in the system, the Callisto system potentially creates two distinguished events to the authority under my two-stage setup: (1) no report (I denote this event by *N*); (2) two reports (I denote this event by *rr*). It is possible that we could have either two immediate reports or a delayed report followed by an immediate report. However, the Callisto system makes those two events the same. To give reporter a strict incentive to report, *rr* should lead to investigation and possibly disciplinary action. No report lead to no action by the authority. This requires $\hat{p}(rr) \geq s$; and moreover $\hat{p}(N) < s$.

Entering a report into Callisto system without seeing by the public right away helps a victim (liar) to incurs her cost immediately, because all the cost of reporting is related to the arrest that follows two reports only. Therefore, A player can never be worse off by sending a report in to Callisto system.

Proposition 4. *Under Callisto system, all victims and liars report right away.* $\alpha(c_v) = \beta(c_l) = 1$ *for any* c_v *or* c_l.

4. Type I and Type II Error

Given that our criminal justice system is imperfect (i.e. the truth state of the world may not always uncover). The authority must strike between the possibility of punishing ***A*** when ***A*** never harassed (type I error) and letting go ***A*** when ***A*** has committed a crime (type II error).

4.1 Type I and Type II Error Outside Callisto system

Recall that in under the normal condition we assume that ***A*** is arrested if and only if there are two reports. (i.e. $\hat{p}(rr, rR) \geq s$). Given every player adopt the equilibrium strategy, the probability that there are two liars and each of them reports immediately is: $(1-p)^2 q^2 b$. The probability that there are two liars, L_2 reports immediately and L_1 delays her report is: $(1-p)^2 q^2 (1-b) b$. Therefore:

$$\Pr\ (\text{Type I error}) = (p-1)^2 bq((1-b)^2 + q)$$

Moreover, ***A*** will not be arrested if we have a single report from either stage or no report (i.e. and $\hat{p}(nn, rn, nr) < s$). Given every player adopt the equilibrium strategy, the probability that V_1 reports and V_2 doesn't is $p^2 a(0)$. The probability that V_2 reports and V_1 doesn't make a delayed report is $p^2(1-a)a(0)$. The probability that neither V_1 nor V_2 reports is $p^2(1-a)^2$. The probability that there is only one victim is $2p(1-p)(1-q)$.
Therefore:

$$\Pr\ (\text{Type II error}) = p^2(1-a)^2 + 2p(1-p)(1-q)$$

4.2 Type I and Type II Error Under Callisto system

Under the Callisto reporting system, ***A*** is arrested if and only if there are two reports. (i.e. $\hat{p}(rr) \geq s$). The probability that the probability that there are two liars reports immediately is $(1-p)^2 q^2 (1)(1)$. Therefore, the probability of type I error is:

$$\Pr\ (\text{Type I error}) = (1-p)^2 q^2$$

Moreover, $\boldsymbol{A}$ will not be arrested if we have a single report from either stage or no report (i.e. and $\hat{p}(N) < s$). The probability that neither V_1 nor V_2 reports is $p^2(0)(0)$. The probability that there is only one victim is $2p(1-p)(1-q)$. Therefore:

$$\Pr\ (\text{Type II error}) = 2p(1-p)(1-q)$$

Proposition 5. *The existence of Callisto reporting system increase the type I error and reduces type II error comparing to the equilibrium outside the Callisto system.*

5. Conclusion

This paper studies both victims and liars' incentive to report sexual harassment in a model modified from the two-stage dynamic model Lee and Sueng developed. I firstly look into the equilibrium strategy of victims and liars outside and under the Callisto system. By changing the basic set up and the assumption, I find similar results as Lee and Sueng found. There is no significant difference between the equilibrium strategy chose by victims and liars. Public (authority)'s investigation is somehow necessary to revel the truth of a sexual harassment case. The online report system Callisto gives everyone a greater incentive to report immediately. However, as opposite to what lee and Sueng find in their paper, I find that The existence of Callisto reporting system increase the type I error and reduces type II error comparing to the equilibrium outside the Callisto system.

References

Ayres, I. and Unkovic, C. (2012). Information Escrow. *Michigan Law Review*. Vol. 134, no. 16. p. 145-196

Benoit, C., Shumka, L., Phillips, R., Kennedy, M. and L. Belle-Isle. (2015). Issue Brief: Sexual Violence Against Women in Canada. *Status of Women Canada*.

Brennan, S. and A. Taylor-Butts. (2008). Sexual assault in Canada, 2004 and 2007. *Canadian Centre for Justice Statistics Profile Series.* No. 19. Statistics Canada Catalogue no. 85F0033M.

Eisenmann, T. (2006). Platform-Mediated Networks: Definitions and Core Concepts 7 discussing Farrell & Saloner, supra no. 48.

Fowler, T. (2014). How the Retracted Rolling Stone Article 'A Rape on Campus' Came to Print. ABC news. Retrieved from: http://abcnews.go.com/2020/deepdive/how-retracted-rolling-stone-article-rape-on-campus-came-print-42701166

Harvey Weinstein sexual abuse allegations. (2017). Wikipedia. Retrieved from: https://en.wikipedia.org/wiki/Harvey_Weinstein_sexual_abuse_allegations

Kaufman, M. (2008). Care of the Adolescent Sexual Harassment Victim. *Pediatrics.* Vol. 122, no. 2. p. 462-470.

Ladd, J. (2016). The reporting system that sexual assault survivors want. TED Talk 2016. Retrieved from: https://www.ted.com/talks/jessica_ladd_the_reporting_system_that_sexual_assault_survivors_want

Lee, F and Sueng, W. (2017). Credibility of Crime Allegation.

Luce, H., Schrager, S. and V. Gilchrist. (2010). Sexual Harassment of Women." *American Family Physician.* Vol. 81, no. 4. p. 489-495.

Sexual Harassment. (2018). Oxford Living Dictionary. Retrieved from: https://en.oxforddictionaries.com/definition/sexual_harassment

Victim of Sexual Violence: Statistics. (2015). The Rape, Abuse & Incest National Network (RAINN). Retrieved from: https://www.rainn.org/statistics/victims-sexual-violence

Victor, D. (2017). How the Harvey Weinstein Story Has Unfolded. The New York Times. Retrieved from: https://www.nytimes.com/2017/10/18/business/harvey-weinstein.html