SEXUAL HARASSMENT POLICY AND INCENTIVES TO SOCIAL INTERACTION

May 12, 2008

Elizabeth Walls

Department of Economics
Stanford University
Stanford, CA 94305
bwalls@stanford.edu

under the direction of
Prof. Timothy Bresnahan

ABSTRACT

Laws discourage both illegal activity and legitimate activity that could be mistaken as illegal, particularly if the penalty is sufficiently high. This paper demonstrates that policies designed to fight sexual harassment can lead to a decrease in appropriate social interaction due to the potential for mistake and the significance of its penalties. The model describes the expected payoffs of communication and how such policy affects them, demonstrating that if the possibility of mistake exists, anti-harassment policy will deter innocent communication. The paper further describes the effects of anti-harassment policy in a hierarchical, two-gender workplace in which social relationships with superiors and colleagues influence whether a worker receives a promotion. The model shows that in this situation, anti-harassment policy primarily harms female employees, and harms them more if they are the minority gender in the workplace. The model holds implications not only for anti-harassment policy but for any area featuring by reporting and the potential for mistake.

Keywords: Sexual harassment, regulation, law, workplace, communication, interaction, gender, discrimination, glass ceiling, Title VII, social capital.

Acknowledgements: Many thanks to Tim Bresnahan for being an advisor in the truest sense of the word, to Geoffrey Rothwell for keeping me on track, to Gavin Wright for introducing me to economics, and to Mark Tendall for helping me learn how to be an economist. Thanks also to Noah Richmond for so much help with the painful early drafts.
INTRODUCTION

Sexual harassment has the rare distinction of being intensely private in nature and extremely public in its ramifications. Companies that would not be held liable if one of their employees punched another employee are held accountable if an employee sexually harasses another. Because matters of sex in the workplace hold a unique position under the law, they have gained unique treatment in the firm as managers have learned to fear the possible lawsuits filed by harassment victims, disgruntled employees, or the Equal Employment Opportunity Commission (EEOC).

Harassment is defined by the EEOC as “unwelcome sexual conduct that is a term or condition of employment,” whether this is a demand for sexual favors or the creation of a hostile work environment. United States laws and policies against sexual harassment are built on the prevention of discrimination; the EEOC first issued guidelines forbidding sexual harassment as sex discrimination in 1980. The Supreme Court upheld this standard in 1986 in Meritor Savings Bank v. Vinson, stating, “A claim of ‘hostile environment’ sexual harassment is a form of sex discrimination that is actionable under Title VII.” Title VII, a provision of the Civil Rights Act of 1964, was the first federal statutory prohibition of discrimination on the basis of sex, and after Meritor, the Supreme Court continued to apply Title VII to harassment cases. In response, firms began thoroughly training their employees not to harass each other in a desperate attempt

---

1 The Civil Rights Act of 1964 was primarily designed to combat racial discrimination, particularly in public accommodations. In an interesting historical accident, the provision against sex discrimination was included by opponents of the Act as an attempt to ensure its defeat.
to reduce their legal liability. The prevention of harassment lawsuits has created its own industry of lawyers, human resources professionals, and consultants.²

Over the past twenty years, a few economists have followed lawyers and managers in paying attention to sexual harassment in the workplace and the policies designed to address it. Apart from the obvious possibility for normative, moral, and emotional arguments on the issue, most economists who have addressed harassment fall into one of two camps. Some believe that policies forcing the reduction of harassment are necessary and efficient;³ others believe that such policies are a superfluous government-induced interference in a problem better solved by natural market correction.⁴

In this debate, little research has focused on the effects of harassment policy on both male and female individuals at the micro level. Such analysis is difficult in respect to harassment because of harassment’s inherently subjective relationship to personal interactions. Determining whether ‘harassment’ occurred requires the imposition of broad standards on situations that are necessarily particular. Legal scholarship has recognized the problem of strict anti-harassment policies because of the possibility of mistake, a problem that arises because harassment is often contextual. Robert Lee and Paul Greenlaw (1995) remark that, particularly because of the 1993 Supreme Court decision Harris v. Forklift Systems, Inc., “…employers can anticipate suits involving situations that may seem inoffensive or reasonable to most of their men workers…”

---
³ See Donohue (1989) and Basu (2003), reviewed in Section 1 of this paper.
⁴ See Posner (1999), reviewed in Section 1 of this paper.
(363). Many firms choose a stricter policy against harassment than is even required by law; Vicki Schultz observes,

A huge (and growing) literature warns companies that they should go beyond the dictates of the law to curtail broad forms of sexual conduct – including conduct that does not satisfy the legal definition of sexual harassment and that does not necessarily undermine gender equality on the job – in order to avoid liability for sexual harassment (2090).

Even employees sensitive to any malicious or egregious harassment may find it difficult to tell when a romantic or quasi-sexual interaction is unacceptable. They may even be unable to determine whether a friendly social interaction might be interpreted as sexual.

This paper will examine the ways an individual’s intra-workplace communication choices are affected by the interaction of harassment policy and the acknowledged possibility of mistake. Through a game-theoretic model of communication, the paper will formalize the notion that if an employee fears being mistaken as a harasser, he has a decreased incentive to interact socially in the workplace. The paper will thus explore the costs of anti-harassment policies incurred by female employees and provide an explanation for a continuing “glass ceiling” preventing women from reaching the highest-ranking positions as easily as their male colleagues. It will further demonstrate the effects that changes in gender composition of the workforce have on the influences of anti-harassment policy and on potential victims of harassment. Broadly, the model holds implications for how firms should shape policies about sexual harassment in order to minimize harassment without obstructing non-harassing communication.

The paper will begin by reviewing the (somewhat sparse) economic and other academic literature related to sexual harassment policy in Section 1. Section 2 will establish and explain some basic assumptions of the model. Section 3 will introduce the model, describing the
decision to communicate and how harassment changes this decision, focusing on interactions between colleagues. Section 4 will adjust this model, applying it to a scenario in which promotion depends on the size of one’s social network and expanding to consider supervisor-employee harassment. Section 5 will discuss the results and conclude.

1. LITERATURE REVIEW

Though there is currently little economic theory on sexual harassment policy and its effect on communication, significant research relates at least partially to the topic. The literature can be divided into five areas: 1) empirical research on the effects of harassment on individuals; 2) empirical research on harassment’s effect on organizations; 3) empirical research on social capital; 4) theoretical examinations of harassment law; and 5) social capital models.

1.1. Empirical research – harassment and individuals

Both psychology and economics inform current empirical research on harassment and the individual. Psychologists have proposed that a woman who experiences sexual harassment in the workplace is measurably harmed by it: increased job stress, a more negative disposition, or poorer psychological health. Economists have tested this hypothesis by analyzing the effect of harassment on variables including job satisfaction, earnings, or an employee’s intention to quit in the near future.

Two such examinations of the topic are by Laband and Lentz (1998) and Antecol and Cobb-Clark (2006). Laband and Lentz explore the interaction between harassment and the three variables listed above for female attorneys in the US, using a survey performed in 1990 by the

---

5 See Schneider, et al. (1997) for one discussion; Antecol and Cobb-Clark (2006) cite several more examples of relevant psychological literature.

Responses to the survey used by Laband and Lentz include information on respondents’ job satisfaction and on whether they had experienced or witnessed sexual harassment by a boss, a colleague, or a client. Antecol and Cobb-Clark’s data set includes information on actual occurrences of conduct that constitutes harassment, separated from whether a person considered herself as having been harassed.

Both papers attempt to control for relevant variables related to professional advancement and job satisfaction to capture the effect of harassment on an employee’s work experience. Laband and Lentz test for the coefficient on experiencing or observing harassment using a probit model, with job satisfaction, earnings, and “quit intention” (a woman’s intention to quit her job within the next two years) as dependent variables in separate regressions. They control for factors such as marital status, prestige of law school, etc. and whether or not a firm falls under Title VII’s prohibition of sex discrimination – i.e. whether the firm had more than ten lawyers. (Under current federal law, small firms may sexually harass their employees freely.)

Like Laband and Lentz, Antecol and Cobb-Clark test for the effects of harassment on job satisfaction and intention to quit (which they call “intended turnover”), though not for its effects on earnings. They use both a single probit model to analyze the effects of harassment on job
satisfaction and a bivariate probit model that includes whether or not a woman felt harassed, as well as the variable on actual harassment.6

While the authors of both of these papers use similar statistical approaches, they reach radically different conclusions. Laband and Lentz find statistically significant results that sexual harassment leads to lower job satisfaction and higher intention to quit (though it has no effect on earnings) and that the magnitude of these effects depends on the source of the harassment. When their model separates harassment by supervisors, colleagues, and clients into separate regressions, the negative effect of colleague harassment is only two-thirds as large as the effect of harassment by supervisors. When all three potential types of harassers are included in a single regression, only harassment by supervisors has a negative effect.

Laband and Lentz differentiate alleged harassment from actual harassment by using the rate at which men describe having observed harassment (relative to how often women describe observing harassment or experiencing it) as a proxy for actual harassment. The authors assert that this approach effectively demonstrates that harassment complaints are grounded in reality because the number of incidences of women experiencing or observing harassment is correlated with the number of incidences of men observing harassment.

Antecol and Cobb-Clark find any negative effect of harassment to be overstated because of this choice of proxy in Laband and Lentz’s model. They claim that Laband and Lentz’s

---

6 Antecol and Cobb-Clark glean their data on actual harassment from survey questions asking whether the respondent has experienced some specific behavior that is considered harassment, such as unwanted propositions for sexual favors or unwanted repeated pressure for dates. Their data on whether the respondent felt harassed comes from direct questions asking whether the respondent has experienced (unspecified) harassment or whether the respondent has felt sexually harassed. Some respondents do answer affirmatively to questions in one of the categories and not in the other.
results are susceptible to a problem of endogeneity with reports of sexual harassment – that women who report sexual harassment may be generally less satisfied with their jobs because of some hidden characteristic. Antecol and Cobb-Clark’s results show that actual harassment does not significantly affect on job satisfaction or intended turnover when the authors use a bivariate probit model to control for the woman’s perception of harassment. They note that their own single-variable model overstates the effect of harassment and that the bivariate probit model prevents such omitted variable bias by controlling for underlying, unobserved personality traits. They find that it does not matter whether a woman experiences behavior that could be generally considered harassment. What does matter is whether the woman feels that she has been harassed. Her perception of the behavior, not the behavior itself, affects her job satisfaction: the authors state, “Women who view their experiences as sexual harassment suffer…consequences over and above those associated with the behavior itself” (56-57). Antecol and Cobb-Clark’s finding calls into question whether sexual harassment is always bad for a harassment victim’s productivity, even if she does not call the behavior harassment. Other economists do not consider this distinction and treat harassment’s negative effect on productivity as a certainty.\footnote{As of March 30, 2008, no published paper in the Social Science Citation Index cites Antecol and Cobb-Clark (2006). It seems that any refutation of either their methods or their conclusions is yet to appear.} For example, building on previous psychological research, Faley, et al. (1999), take the “broad array of negative personal and professional consequences” (462) of harassment as given.

Economists have also examined individual perceptions of harassment. Antecol and Cobb-Clark (2003) explore the effects of sexual harassment training on whether employees perceive certain behaviors as harassment. They use data from the Merit Systems Protection
Board of the US federal government (USMSPB) collected in 1994 from federal employees. Employees answered whether they considered four types of unwanted sexual behavior to be harassment if the behavior came from a supervisor and separately responded to questions on the same behaviors from co-workers. Employees were also asked about previous harassment training and whether they thought it had affected them and their organization.

Controlling for relevant demographic variables, the authors use a probit model in which the dependent variable is the likelihood that an employee considers one of the subcategories of behavior to be harassment. They find a significant, positive coefficient on whether the employee has had training about sexual harassment and that training has a larger effect on men’s responses than on women’s. They find both genders are more likely to perceive a given unwelcome action from a supervisor to be “harassment” than they do the same action from a colleague. Antecol and Cobb-Clark also observe that if someone has had harassment training, they are more likely to view both unwanted supervisor conduct and unwanted colleague conduct as harassment, but the magnitude of the effect of training is larger regarding perceptions of colleague harassment.

The authors’ results lead them to approve of such harassment training, describing its effects positively as “especially helpful in clarifying men’s views about unwanted sexual behavior” (841, emphasis added).

Antecol and Cobb-Clark’s conclusions on the potential mutability of harassment perceptions relates directly to their findings on harassment in the armed forces. If individuals who have received harassment training are more likely to perceive actions as harassment, and if job satisfaction is affected by perception of harassment rather than actual harassment (as they
find in 2006), harassment training may have the problematic consequence of increasing job
dissatisfaction. Such a result calls into question the value that Antecol and Cobb-Clark perceive
in harassment training, at least for women as the most likely potential victims of harassment.

These empirical results indicate the need for greater study of harassment by colleagues,
which has largely received parenthetical dismissals by the few economists to consider
harassment theoretically.\textsuperscript{8} Intuitively, there is a greater chance of harassment between two
equals is more likely to occur, as co-workers interact with each other more frequently than with
their supervisor. Even if a given worker interacts with her supervisor as frequently as with other
colleagues, she probably has more peers than supervisors.

This intuition is backed by evidence; based on the USMSPB data, Antecol and Cobb-
Clark (2003) note in an aside, “...most harassment occurs between co-workers rather than
between supervisors and subordinates” (841). This finding holds relevance when viewed
alongside Laband and Lentz’s finding (discussed earlier) that harassment by colleagues does not
seem to have an effect on employees that is comparable to the effect of harassment by
supervisors (1998). These two findings, if both true, suggest that the gains from training,
policies, or legal rules focused on eliminating colleague harassment may be small.

Though not directly related to harassment, Rees et al. (2003) find that individual
productivities are interdependent within a team and establish such interdependencies even in the
absence of group-based output or technological dependence. They show that one source of this
phenomenon is the exchange of information, which could be inhibited by overly strict anti-

\textsuperscript{8} See Basu (2003), Donohue (1989), and Posner (1999).
harassment regimes. The authors find that productivities are particularly interdependent for salaried workers, as these categories can have weak individual incentives to perform. This result supports the importance of examining any policy with the potential to reduce communication.

1.2 Empirical research – harassment and the Army

In addition to researching the effects of harassment on individuals within a firm, economists have also examined aggregate consequences to organizations. Faley, et al. (1999) estimate the cost of sexual harassment to the entire US Army for the year 1988 – though, notably, they assume that harassment creates significant costs for individual women without considering the potential for endogeneity raised by Antecol and Cobb-Clark (2006). In the estimate, the authors divide costs into three categories, each with subcategories. The first category is productivity-related costs: subcategories are lost productivity, incident costs, and costs of increased absenteeism. The second category is administrative costs: subcategories are costs from separation, replacement, and transfer. The third category is “other” costs: legal, medical, and counseling costs. They estimate the costs of each of the nine subcategories individually and aggregate them. The final estimate is that the cost of sexual harassment to the Army in 1988 alone was $250 million (in 1994 dollars).

The paper does not address why, if harassment is so expensive, the Army allowed such unnecessary costs to accrue. They do not consider the potential difficulties and costs in decreasing harassment – costs that may be large enough to dissuade a profit-maximizing entity from attempting to stop harassment within the firm.
1.3 Empirical research – social capital

Economists give varying definitions for the term “social capital,”\(^9\) but the term generally refers to the value or productive value of connections among people and groups of people. As social capital relates closely to issues of interaction amongst employees, job satisfaction, and productivity determinants, recent research on the topic hints at a cost that firms might incur by either enduring or removing harassment. Helliwell and Huang (2005) consider the problem of omitted variable bias in the use of hedonic price methods to determine how much people value non-financial work factors. They claim that because of unobserved differences in talent, the value of these non-financial factors will be understated by any method that compares market wage rates.

Their model evaluates how much a given job characteristic affects “life satisfaction,” which is measured using survey responses on well-being. They also evaluate the extent to which income affects life satisfaction, which allows them to monetize the value of job characteristics without using wage differentials. Helliwell and Huang use data from three separate Canadian surveys with data on life satisfaction, personality traits, demographics, and jobs.

They find that measures of workplace social capital significantly affect life satisfaction. Specifically, they find that of all measures of social capital, workplace trust influences satisfaction most strongly. They also show that this result is not a function of unobserved personality traits, as they control for these as well. Even after controlling for whether a person is positive, friendly, or some other characteristic that one might imagine being associated with

higher life satisfaction, trust in co-workers still has a significant and large coefficient. Helliwell and Huang calculate (based on their estimate of income’s effect on life satisfaction) that the value to a worker of adding one point on a ten-point scale ranking trust in co-workers is worth $13,000 (relative to a modal income of respondents of $65,000). This result suggests the importance of examining any policy that requires employees to report their colleagues, which may weaken workplace trust.

Based on these results, Helliwell and Huang conclude that there are unexploited gains to be had for both employers and employees if they can encourage trust in the workplace. However, they do not explain why these potentially large gains remain unexploited, nor do they consider the difficulties in creating “trust.” This finding does indicate that workers value social capital in the workplace, supporting the idea that a policy that decreases social capital causes a loss for firms and their employees. Helliwell (2004) supports the general validity of social capital; in this work, he finds that across countries, national social capital levels were strongly correlated with national suicide rates. Helliwell claims this indicates that subjective measures of social capital authentically represent real human happiness.

1.4 Theory on harassment law and banning harassment

Regardless of any costs to women, the prohibition of workplace harassment poses a theoretical difficulty for a classical economist. A ban on harassment could be considered a ban on a particular kind of voluntary contract, and in theory, voluntary contracts enhance efficiency. An employer who wishes to harass his female employees could pay women a higher wage in return for being allowed to harass them. If the wage differential was larger than the value of the
discomfort the women experienced, they would be happier in such a role. Donohue (1989) cites an earlier paper by Posner as opposing a ban on harassment on these grounds.

Basu (2003) creates a model to demonstrate why it is not efficient for women to accept jobs at firms known for sexually harassing people, even if they know this fact ahead of time. Basu’s model first describes a world in which harassment is legal as long as it is disclosed at the time of the labor contract. His model throughout the paper allows only for employer-employee harassment, not harassment by peers, and assumes that harassment is binary (it either occurs or it doesn’t). Because the employee is informed before s/he chooses whether to form a contract, a compensating differential emerges and two wage rates prevail: a non-harassment wage ($w_N$) and a harassment wage ($w_H$). Basu then models a world in which harassment is totally banned, which is a simple labor demand-supply model resulting in one wage rate, $w^*$.

Donohue’s and Basu’s models both find that a ban on harassment is efficient, but on different grounds. Basu’s model illustrates that in a world without harassment, $w^*$ is between $w_N$ and $w_H$. Based on this, Basu argues that everyone who chooses a job that promises no harassment is better off when the law prohibits harassment, as they all receive $w^*$ instead of $w_N$. Though such a result could apply to any workplace characteristic, Basu uses this premise to conclude that there is no efficiency argument against a ban on harassment. He permits himself to make normative assessments of regimes permitting harassment; most notably, he claims that the right to be free of it is an inviolable right.

In contrast, Donohue’s model concedes the ‘inefficiency’ due to forbidding such contracts in a static supply-and-demand framework. However, he attacks the idea that
harassment should be permitted by citing evidence showing that such preferences are “malleable” and can be removed (1339). If these preferences can be removed by the imposition of laws against harassment, the effect would be the same as any other technological advance that generated the same output for lower labor or capital costs. Donohue asserts that harassment law is therefore dynamically efficient.

Donohue does not adopt Basu’s premise of a situation in which harassment is legal as long as female employees know about it beforehand, as this scenario has never existed in the United States. Donohue instead claims that information asymmetries are a reason harassment cannot self-correct: when a woman takes a job, she does not know whether she will be harassed, so she cannot refuse the job because of it or demand higher wages in exchange for the harassment. In addition, he asserts that because victims of harassment have no incentive to furnish information about this harassment, employers and potential employees do not receive any signal about the harassment. Donohue uses this as another reason for a ban on harassment.

Basu concludes that harassment should be banned because it results in lower wages for workers who wish to avoid harassment and that some people who would be happy earning $w^*$ in a harassment-free society must (in order to work) take a higher-paying, harassing job if harassment is allowed. In contrast, Donohue’s conclusions are driven by his idea of dynamic efficiency. Donohue supports his claims to the benefit of banning harassment by providing statistics showing that between 1966 and 1977, female productivity rose relative to the productivity of men. He also shows that the female labor supply curve shifted out after 1964. Though he controls his productivity measure by comparing female productivity growth to a
baseline of male productivity, Donohue acknowledges that many other factors may have affected these changes as it impossible to control for everything that may have influenced productivity and gender relations.

Economists have also explored how harassment might be eliminated in addition to whether or not it ought to be. Posner (1999) ignores his former objections to eliminating harassment and instead opposes the current legal structure of harassment law. His primary objection to the law is that it is unnecessary. He asserts that if firms are harmed by harassment, since all parties concerned are employed within the firm, they should be able to reach the optimal outcome without external interference like the ban Donohue and Basu support. Because the dispute is within the firm, he believes that firms have the ability to reach a perfect Coasian solution – a notion that leads him to skepticism on the validity of sexual harassment as a tort. Posner further believes that the market is capable of disciplining less efficient producers, including harassing firms. Donohue counters this idea by claiming the law will remove discriminators faster than the market and that society thus benefits by more quickly eliminating discriminators and the deadweight loss they cause.

Posner also claims that harassment law creates unnecessarily large enforcement costs, as the government forces employers to be more vigilant than is efficient. Posner further declares that because of these costs, harassment law hurts average people, including most women. However, Posner provides no actual estimate of enforcement costs; Donohue estimates these cost to be only about $8 per employed woman. Further, Posner does not provide actual evidence that women have been hired less frequently or paid less because of harassment law.
Of these economists, only Posner considers the possibility of a decline in social relationships, an idea he considers only in passing. He mentions that many people search for spouses in the workplace and that a ban on intra-office relationships would harm these employees. Posner does not consider the effect on work activities of eliminating relationships.

The economic research also investigates topics relevant to sexual harassment law that are not explicitly related. Athey, et al. (2000) create a model of firm diversity in which upper-level employee diversity is endogenously determined by the previous amount of upper-level employees of a certain type. In their model, employees pass down human capital through mentoring and such human capital flows more easily between two employees if they are the same type, making a firm better off in the long run to have a diverse workforce under certain conditions. Their model holds implications for the value of diversity and the potential obstructions to diversity that can result by impeding communication (and “mentoring”) between two different types – or genders, as in the case at hand.

1.5 Social capital models

Spagnolo (1999) creates a theoretical framework for exploring workplace social capital that focuses on its value to firms, rather than to individuals. He creates a game-theoretic model of workplace cooperation based on social capital “linkages,” or connections between production relationships and social relationships. He illustrates the ways in which a strong social relationship can use its influence to create cooperation in production if the two relationships affect each other.
His model uses a two-person prisoner’s dilemma game in which total gains are higher if both parties cooperate but in which both players have incentives to deviate. The game of social interaction is separate from the game of workplace interaction, but the two can be linked if the agents in the two games are the same or if the outcome in one determines the outcome in the other. Spagnolo assumes that social capital allows players to enforce cooperation. He also assumes that the payoffs from either social or production relations are substitutes, an intuition which is supported by Helliwell and Huang’s findings. If this assumption holds, Spagnolo’s model predicts that linking the two relations will always facilitate cooperation in both – because it increases the benefits of cooperation (or increases the harm from non-cooperation) and decreases the relative incentive to “cheat.” Further, the presence of social capital can reduce the risk that employees take by cooperating. The model formally demonstrates the intuition that if you trust someone away from work, you are more likely to trust her at work.

Spagnolo’s most pertinent conclusion is that social capital can be used to increase cooperation in the workplace. His results also indicate that changes in either a social relationship or a production relationship can lead to escalating effects in both if one starts to fall apart. His model leads him to believe that firms have incentives to employ people with existing social capital (perhaps from the same area, culture, ethnic group, etc.) and to encourage social interaction among employees. Such conclusions are consistent with Athey, et al.’s model in which a mentor creates human capital in the mentee.

Spagnolo further proposes his model as a possible explanation for occasional “irrational” behavior by employees towards their co-workers. He also mentions the counterintuitive
possibility that a firm may induce costs by weakening a worker’s union, because lost social capital between the employees might decrease their ability to cooperate in the workplace. Spagnolo’s model is unique in its focus on social capital in the workplace and provides possibilities for potential extrapolation to a workplace altered by sexual harassment law. His model demonstrates that firms, as well as individuals, benefit from social relationships with their co-workers.

2. UNDERLYING ASSUMPTIONS

This paper will focus on harassment occurring through direct and intended communication between individuals. Examples of such harassment include off-color remarks, inappropriate jokes, and propositions. For simplicity, the paper will not examine cases in which there is no intention to interact, or in which the intent to interact is less clear. An example of this type of harassment is a situation in which one person displays an image at his/her desk that offends another person who sees the image. The action of displaying an image cannot entirely be defined as a way of communicating with another person, as the person who does so may or may not have any intention for others to be aware of the action and the image.

When this paper discusses social interactions, the term is intended rather literally; it includes any interaction that is not strictly required for the completion of an individual’s job. For simplicity, we further assume individuals are risk-neutral regarding the payoff of such interactions.

For the purpose of this study, all harassment will be assumed to be male harassment of females. In reality, this is not universally the case – the USMSPB reported that 19% of male
federal employees who responded to their survey had experienced some type of harassing behavior in the past two years (13). Further, the Supreme Court has ruled in Oncale v. Sundowner Offshore Services, Inc. (1998) that the respective gender of the harasser or harassee does not preclude sexual harassment. However, most harassment originates with men and is directed towards women; the USMSPB found that 44% of female federal employees who responded to the survey had experienced harassment and that “the overwhelming majority of females (93 percent) have been harassed by men” (18). Thus, this study will use the narrowing assumption that men harass women, though the structure of the model still holds for any potential harasser or potential harassee regardless of gender.

Under the law, harassment comes in two forms – quid pro quo harassment and hostile work environment harassment. However, in an economic context, the distinction between the two types of harassment is relevant only insofar as it reflects the identity of the harassing party, rather than the precise nature of the harassment. Quid pro quo harassment is always harassment originating with a person who has power over the harassee; it is impossible to classify the harassment of an equal as quid pro quo because there is no career-oriented gain or punishment for refusal or reward. Thus, the law would classify any harassment from a peer as hostile work environment harassment.

Similarly, any hostile work environment harassment from a supervisor can be viewed as a type of quid pro quo harassment: a victim of a hostile environment of a supervisor’s creation may receive the spoken or unspoken message that she risks her future success within the firm if

---

10 In Oncale, the Court found that workplace harassment can violate Title VII even if the harasser and harassed employee are the same gender.
she does not accept the environment. Thus, in this paper, harassment will be considered not on the basis of its legal classification but rather on the identity of the harasser.

Any further assumptions related to specific parts of the model will be described in the relevant section. Section 3 will focus on harassment between colleagues, cases in which no party has power over the other. Section 4 will then extend the model, applying it to a promotion scheme in which some individuals are superiors and have the power to decide which outsiders are promoted. Section 5 will conclude.

3. A BASIC MODEL OF COMMUNICATION

3.1 Model of communication between employees

In this model, there are both male and female workers. There are two types of male workers: those who do not receive any personal benefit from harassing someone (Type G) and those who do benefit from or enjoy harassing (Type H). Retaining the simplifying assumption that people cannot harass others of the same gender, Person 1’s expected payoff from communicating with a coworker of the same gender is given by the following equation, which forms the baseline of the communication model:

$$E[Payoff_{f_1}(X)] = p_g(X_g) + p_i(0) + p_b(X_b) \quad \text{s.t. } p_g + p_i + p_b = 1$$

- \(X\) is the social interaction
- \(X_g\) is the positive result from a good social interaction
- \(X_b\) is the negative result from a bad social interaction
- \(p_g\) is the probability of a successful interaction (which results in \(X_g\))
- \(p_i\) is the probability that the partner is indifferent to the interaction (no benefit or harm)
- \(p_b\) is the probability of a bad interaction (which results in \(X_b\))

All the interactions in the above case are intended as non-harassing social interactions. The probabilities of various outcomes are determined by the responses of the partner, Person 2 – s/he
may welcome, ignore, or rebuff the interaction. Person 1 determines for himself or herself the value, or payoff, of each kind of interaction (represented in Equation 1 by $X_g$ and $X_b$). For simplicity’s sake, we assume that Person 1 knows (or at least can estimate) the values of the probabilities, which are determined by Person 2, and can thus determine an expected payoff of the interaction.\footnote{This may not be precisely true in all cases, but most people can estimate with some degree of accuracy the reception their interaction is likely to receive, whether they base this guess on past interactions with the second party or on their own previous interactions in general.} Because Person 2 determines the probabilities and Person 1 chooses the values of each type of outcome, the expected payoff of the interaction for Person 1 is determined by both parties.

The value that an individual assigns to a “good” or “bad” interaction need not be constrained relative to the other. For example, someone might choose not to interact even if the probability of a good interaction is high. If Dan actively dislikes Ethan, Dan might assign a lower or even negative score to a “good” interaction with Ethan because he would prefer a “bad” one – Dan might prefer to undergo some unpleasantness rather than have a friendly conversation with Ethan.

\subsection*{3.2 Communication and the possibility of harassment: female-initiated interactions}

If a female cannot distinguish whether a man is Type G or Type H, her expected payoff to interacting with any man is based on the overall probability she expects of harassment:

\[ E[Payoff_f(X)] = (p_g - h)(X_g) + p_l(0) + (p_b + h)(X_b) + h\text{Harassment} \]  \hspace{1cm} (2)

$h$ is the probability that the female will experience what she interprets as harassment during the interaction.

$0 \leq h \leq 1$

$h < p_g$ else a minimum value of 0 replaces the term $(p_g - h_1)$
\( h < 1 - p_b \) else a maximum value of 1 replaces the term \((p_b + h_1)\).

Harassment is the disutility a woman experiences from undergoing what she perceives as harassment; \( \text{Harassment} \leq 0 \).

\( Xs \) and \( ps \) are as in Equation 1.

This modification of Equation 1 represents that the probability of a good interaction is decreased by the probability of harassment \((h)\) and the probability of a bad interaction is increased by that amount. In addition, the negative effect of harassment itself is included as the term \( \text{Harassment} \).

If the woman could discern a man’s type, she would generally be more likely to interact with a Type G man because there is a lower probability that she will experience harassment from him. Though the Type G man receives no benefit from harassment (by assumption), the probability is nonzero because of the potential for mistake. A woman could still feel that her interaction with a Type G man included harassment without him intending this. Her interaction with a Type H man would have a higher, though not 100\%, chance of incurring harassment. Her interactions with a Type H man and a Type G man are, respectively, as follows:

\[
E[\text{Payoff}_1(X^H)] = (p_g - h_a)(X_g) + p_l(0) + (p_b + h_1)(X_b) + h\text{Harassment} \\
E[\text{Payoff}_1(X^G)] = (p_g - h_b)(X_g) + p_l(0) + (p_b + h_2)(X_b) + h\text{Harassment}
\]

\( h_a \) is the probability that the female will experience harassment during her interaction with the Type H male.

\( h_b \) is the probability that the female will experience harassment during her interaction with the Type G male.

\( h_b < h_a \)

Since \( h_b < h_a \), her expected payoff from the interaction is higher with the Type G man because of the decreased probability of a bad interaction and of harassment.

Whether a female can distinguish between the types or not, these equations indicate the potential gains to female employees from deterring harassment. Were it eliminated, she would
no longer face the negative consequences of harassment and she would be more likely to interact with her male coworkers, increasing her social capital. The model demonstrates this plainly: returning to the assumption that the female cannot distinguish between the two types, the change in her expected payoff as harassment increases is given by the derivative of Equation 2 with respect to $h$:

$$\frac{dE}{dh} = -X_g + X_b + \text{Harassment}$$  \hspace{1cm} (5)

Since $X_g$ is generally positive, $X_b$ is generally negative, and Harassment is negative, the change with respect to $h$ will be negative. If a woman is more likely to expect (what she perceives as) harassment, she is less likely to initiate an interaction.

3.3 Communication and the possibility of harassment: interactions initiated by Type H males

When a man initiates contact with a woman, he can initiate either a harassing interaction or a non-harassing interaction. Either type of man (G or H) can initiate either type of interaction. If a Type H man initiates a harassing interaction, his expected payoff is:

$$E[Payoff_1(X)] = (p_g - h_1)(X_g) + p_i(0) + (p_b + h_1)(X_b) + \text{Satisfactions} + a \ast h_1(\text{Consequences})$$  \hspace{1cm} (6)

$h_1$ is the probability that the female recipient of the interaction will interpret the interaction as harassment.

$a \ast h_1$ is the probability that the female will report the interaction as harassment.

$0 \leq a \leq 1$

Satisfactions is the benefit that the man receives from the harassing interaction; for a Type H man, Satisfactions > 0.

Consequences represents the negative ramifications such as investigation, litigation, losing a job, etc. that may result from being reported for harassment.

$Xs$ and $ps$ are as in Equations 1-3.
In this scenario, *Consequences* can represent legal penalties or merely penalties imposed by the firm itself; any negative result a man risks incurring when he harasses. In a world in which harassment is permitted, *Consequences* would be 0.

The probability of a “good” interaction does not automatically go to 0 solely because the interaction is harassing and $h_1$ need not be 100%: there exists some possibility, however small, that a woman would not interpret the harassment as such or that she might not mind it. In this equation, $h$ is multiplied by the term $a$ because even if a woman interprets an interaction as harassment, she will not always report the harassment. Schneider, Swan, and Fitzgerald state, “Few… women confront the harasser, and very few women seek formal relief from the organization” (403). Their findings imply a low value of $a$ and support the fact that not all women who perceive harassment report harassment.

The payoff is the same as Equation 6 for a Type G man who initiates a harassing interaction as for a Type H man, except for the personal benefit that each type receives from harassing. The distinguishing characteristic between the two is that for the Type G man, $Satisfactions \leq 0$; for the Type H man, $Satisfactions > 0$. Nothing else in the harassment equation changes, including the parameter $h_1$. If $Satisfactions < 0$ for the Type G man, his equation would be the same as for a Type H man (though leading to different results). If $Satisfactions = 0$ for a Type G man, the simpler expression for his expected payoff of a harassing interaction is:

\[ E[Payoff_f(X)] = (p_g - h_1)(X_g) + p_i(0) + (p_b + h_1)(X_b) + a \cdot h_1(Consequences) \] (7)
For a non-harassing interaction, the expected payoff for a man of either type (G or H) is as follows:

$$E[Payoff(X)] = (p_g - h_2)(X_g) + p_l(0) + (p_b + h_2)(X_b) + a * h_2(Consequences)$$

(8)

$$h_2$$ is the probability that the recipient of the interaction will interpret the interaction as harassment (even though it is not intended as harassment).

$$h_2 < h_1$$

A Type G man will choose the interaction that leads to the highest expected payoff. The difference between the two payoffs (of the harassing interaction and the non-harassing interaction) is expressed by the difference between Equation 7 and Equation 8, which is equal to:

$$(h_1 - h_2)(X_g + X_b + a * Consequences)$$

(9)

The difference between the two interactions is the difference between the two probabilities multiplied by the values assigned to the various outcomes. This expression gives the minimum difference between the two payoffs, as it assumes that the Type G man does not assign a negative value to a harassing interaction (which would represent active distaste for harassing, rather than mere indifference to it). Regardless, because $$h_2 < h_1$$, a Type G man will always prefer a non-harassing interaction.

Recalling that we assume risk neutrality, a Type H man will be indifferent when choosing between a harassing interaction and a non-harassing interaction if his expected payoff is the same for each case, i.e. when Equation 7 equals Equation 8:

$$E[Payoff(X)] = (p_g - h_1)(X_g) + p_l(0) + (p_b + h_1)(X_b) + Satisfactions + a * h_1(Consequences)$$

$$E[Payoff(X)] = (p_g - h_2)(X_g) + p_l(0) + (p_b + h_2)(X_b) + a * h_2(Consequences)$$

$$Satisfactions = (h_1 - h_2)(X_g - X_b - aConsequences)$$

(10)
This illustrates that the benefit the Type H man receives from harassing (Satisfactions) must be greater than the change in probabilities multiplied by the difference between a good outcome and a bad outcome and the chance of Consequences. We can rearrange Equation 10 to get an equation representing the benefit to a Type H man of not harassing, shown in Equation 11.

\[ f = (h_1 - h_2)(X_g - X_b - a\text{Consequences}) - \text{Satisfactions} \]  

(11)

To examine how various factors affect the Type H’s relative incentive not to harass, we take first-order conditions with respect to \( h_1, a, \) and Consequences:

\[ \frac{df}{dh_1} = X_g - X_b - a\text{Consequences} \]  

(12)

An increase in the probability a female will view the harassment as such will increase the Type H male’s incentive not to harass. The higher the probability she will view the action as harassment, the higher the probability she will report it as such and the male will incur consequences.

\[ \frac{df}{d\text{Consequences}} = (h_2 - h_1)a \]  

(13)

Similarly, by increasing the consequences of being reported for harassment, a Type H male is, unsurprisingly, less likely to harass. However, the magnitude of the effect is determined by the difference between \( h_2 \) and \( h_1 \), which is the difference between the probability of a false positive and the probability of a true positive.

\[ \frac{df}{da} = (h_2 - h_1)\text{Consequences} \]  

(14)

\( h_2 \) is greater than \( h_1 \) and Consequences is a negative number, so the term \([(h_2-h_1)\times\text{Consequences}]\) will be a positive number. An increase in the probability that a female will report
harassment when she perceives it as such increases the relative benefit to a Type H man of not harassing – or, more plainly, creates a disincentive to harass.

\[
\frac{df}{dh_2} = -X_g + X_b + a \ast Consequences
\]  

(15)

The crucial result seen in Equation 15 shows that as \( h_2 \) (the probability that a non-harassing interaction will be seen as harassment) increases, the relative benefit of a non-harassing interaction to a harassing one decreases. Because \( h_2 \) is the probability of mistake as it rises, a Type H man has a greater incentive to harass rather than otherwise. If an innocent interaction will be perceived as harassment anyway, the relative value of the non-harassing interaction falls. If, for example, \( h_1 = h_2 \), the Type H man will always choose to harass because he could be reported and face consequences in either case, but only the harassing interaction will give him the enjoyment that he receives from harassment.

3.4 Communication and the possibility of harassment: interactions initiated by Type G males

A Type G man will decide to interact with a woman if the expected value of the interaction is greater than zero. Recall that his expected payoff was described by Equation 8, repeated below:

\[
E[\text{Payoff}(X)] = (p_g - h_2)(X_g) + p_1(0) + (p_b + h_2)(X_b) + a \ast h_2(\text{Consequences})
\]

As \( h_2 \) (the probability that his interaction will be mistaken for harassment) increases, his expected payoff changes as follows:

\[
\frac{\partial E}{\partial h_2} = -X_g + X_b + a \ast Consequences
\]  

(16)
All three terms in the expression are negative numbers; as the probability that a woman will view the innocent interaction as harassment increases, the expected payoff of interacting decreases. As the probability that any perceived harassment will be reported increases, the change in his expected payoff is described by the following:

$$\frac{dE}{da} = h_2(\text{Consequences})$$  \hspace{1cm} (17)

Again, because \text{Consequences} is a negative number, as \( a \) increases, the expected payoff of an interaction decreases. Lastly, a change in \text{Consequences} affects the payoff as follows:

$$\frac{dE}{d\text{Consequences}} = (h_2)a$$  \hspace{1cm} (18)

The effect that a change in \text{Consequences} has on the expected payoff of the interaction is driven by the probability of mistake and probability of perceived harassment being reported.

Thus, increasing the probability that someone who perceives harassment will report it \( (a) \) or increasing the penalties to being reported \( (\text{Consequences}) \) not only deters harassing behavior, as found in Section 3.3, but innocent interaction as well. The magnitudes of the effects are determined by the magnitude of the other determinants: severe consequences lead to a large change in the payoff of the interaction when the probability of reporting \( (a) \) changes, and vice-versa. As the probability of mistake increases, the magnitude of the decrease in the expected payoff is driven by both \( a \) and \text{Consequences}.

The main result of this part of the model is that anti-harassment policy decreases a Type G male’s incentive to interact; fewer social interactions will occur between men and women. This result is driven by the nonzero probability of mistake: if \( h_2 \) were zero, the change in the expected payoff to an interaction would not be affected by reporting of or consequences for
harassment. The chance of mistake creates the deterrence of harmless inter-gender social interaction. A policy that created heavy consequences could deter harassing interactions but also could seriously deter innocent ones, as well. In addition, a policy that seeks to increase reporting of harassment similarly deters both kinds of communication if there is ambiguity as to what interactions will be reported. The model thus demonstrates the challenge faced by a firm in balancing anti-harassment measures while still encouraging interaction, a challenge which will be analyzed further in Section 4 after some additional adjustments to the model.

3.5 Model with exogenous probability of reported harassment

This model has thus far assumed that a female employee is more likely to interpret an interaction as harassment if the action is intended as harassment (in this model, this is given because $h_1 > h_2$). While this seems probable intuitively, as Antecol and Cobb-Clark (2006) found, whether an interaction is perceived as harassment can be affected by exogenous factors. In their case, they discovered that sexual harassment training made people more likely to perceive a given interaction as harassment.

In the case that the probability ($h$) of viewing an interaction as harassment is exogenous (briefly introduced at the end of Section 3.3), we can examine the Nash equilibrium for a Type H man within the model.

<table>
<thead>
<tr>
<th></th>
<th>Call Harassment</th>
<th>Don’t Call Harassment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Female Employee</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Type H Male Employee</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harass</td>
<td>$E(X) + \text{Consequences} + \text{Satisfactions}, E(X) + \text{Net Costs}$</td>
<td>$E(X) + \text{Satisfactions}, E(X)$</td>
</tr>
<tr>
<td>Don’t Harass</td>
<td>$E(X) + \text{Consequences}, E(X) + \text{Net Costs}$</td>
<td>$E(X), E(X)$</td>
</tr>
</tbody>
</table>
If \( h \) is determined exogenously, a Type H man will choose to harass women because he gains benefit from it and, by construction, it has no effect on whether or not the woman reports him. The Nash equilibrium is that the male will harass and the female will not report the harassment. This result is determined by the presence of net costs for a woman to report harassment, such as shame, embarrassment, and time, a notion which is supported by the psychology literature; Schneider, Swan, and Fitzgerald (1997) find that women tend to avoid reporting harassment (404).

If instead there were net benefits to a woman of reporting harassment, the outcome would be as follows:

<table>
<thead>
<tr>
<th></th>
<th>Female Employee</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male Employee</strong></td>
<td></td>
</tr>
<tr>
<td>Harass</td>
<td>Call Harassment</td>
</tr>
<tr>
<td>E(X) + Consequences + Satisfactions, E(X) + Net Benefits</td>
<td>E(X) + Satisfactions, E(X)</td>
</tr>
<tr>
<td>E(X) + Consequences, E(X) + Net Benefits</td>
<td>E(X), E(X)</td>
</tr>
</tbody>
</table>

Here, the Nash equilibrium is that male employees always choose to harass and females always report them. In either case, if \( h \) is exogenous to actual harassment, the dominant strategy for a Type H male is to harass. Thus, if a firm sets or influences \( h \) through its culture or reporting system and does so such that \( h \) is exogenous to actual harassment, they will be incenting rather than deterring such behavior from potential harassers. A practical example of such a case would be if a firm prohibited all romantic or flirtatious interactions between employees and created a policy that any such interactions would be treated as harassment. Such a policy would not deter
harassment from Type H males and would instead deter acceptable, non-harassing behavior from Type G males.

4. HIERARCHY AND COMMUNICATION – EXTENSION AND ANALYSIS

The previous model describes people engaging in isolated bilateral interactions with no payoffs for interaction except utility and potential penalties for harassment. This is not the case in most workplaces, where interactions often determine career advancement and success, as Athey, et al. (2000) demonstrated in regards to the benefits of mentoring. Social interactions matter not both in themselves and because of the social network they generate.

Consider a workplace in which there are two tiers of employees: supervisors and regular employees. In a given period, a regular employee’s chances of being promoted to supervisor depend on the size of the employee’s total social network, the size of the employee’s social network of supervisors, and his/her native ability:

\[ P_1 = \frac{T}{\text{# of employees}} + \frac{S}{\text{# of supervisors}} + \text{ability} \]  

(19)

- \( P_1 \) is the probability that Employee 1 receives a promotion
- \( T \) is the total number of employees in Employee 1’s social network
- \( S \) is the number of supervisors in Employee 1’s social network
- \( \text{ability} \) is Employee 1’s native ability

For simplicity’s sake, let us consider a firm of 76 regular employees and 24 supervisors as an example. The probability that an employee will be promoted is given by the following expression:

\[ P_1 = \frac{T}{100} + \frac{S}{24} + \text{ability} \]  

(20)

or, rearranged:
\[ P_1 = \frac{R}{100} + \frac{S}{100} + \frac{S}{24} + ability = .01(R) + .051\bar{S}(S) + ability \quad (21) \]

Thus, the partial derivatives with respect to the various size of the different networks is simply the constant on the variable:

\[
\frac{\partial P}{\partial R} = .01 \quad (22)
\]

\[
\frac{\partial P}{\partial S} = .051\bar{S} \quad (23)
\]

We now apply the basic bilateral communication model to the intra-firm social network created by these conditions, adding the following simplifying assumptions:

- A successful (“good”) social interaction with a person adds a point value of 1 to the relationship between the people.
- An “indifferent” (or otherwise negligible) interaction adds a point value of 0.
- A “bad” interaction adds -1.
- A person is “added” to a social network when the point values of the interactions sum to 1 as a total.\(^{12}\)
- A potential promotion to supervisor has the same value for all regular employees.
- Non-promotion-related utility of a certain kind of interaction is the same across all other employees, e.g. ignoring the promotion, Employee 1’s values a “good” interaction with Employee 3 the same as with Employee 57.

In addition, we add two assumptions regarding the cost of interaction, which includes time, emotional expenditure, etc. For a regular employee to interact with another regular employee, the cost is 1 unit. To interact with a supervisor, the cost is 3 units.

### 4.1 A single-gender workplace

In a single-gender workplace, free of the possibility of harassment, an employee Kurt can rank various other employees based on his own expected payoff of an interaction, due to the

\(^{12}\) All particular numbers are chosen for perspicuity and could be extrapolated to any other constants.
heterogeneity of the probabilities \((p_g, p_i, \text{ and } p_b)\) of an interaction’s success. In all cases, his payoff equation is, as in Equation 1:

\[
E[\text{Payoff}_f(X)] = p_g(X_g) + p_i(0) + p_b(X_b) \tag{24}
\]

The specific values of the \(X\)s are no longer simply utility-based. Because Kurt is risk-neutral and values the promotion at \(\pi\), the value of his \(X\)s are as follows:

\[
X^R_g = U_g + .01(\pi) \tag{25}
\]

\[
X^R_b = U_b - .01(\pi) \tag{26}
\]

\[
X^S_g = U_g + .051\bar{\delta}(\pi) \tag{27}
\]

\[
X^S_b = U_b - .051\bar{\delta}(\pi) \tag{28}
\]

\(X^R\) is an interaction with a regular employee
\(X^S\) is an interaction with a supervisor
\(U_g\) is the non-promotion-related happiness that Kurt receives from a good interaction
\(U_b\) is the non-promotion-related unhappiness that Kurt receives from a bad interaction
\(\pi\) is the value of the promotion

His expected payoff is then:

\[
E[\text{Payoff}(X^R)] = p_g(U_g + .01(\pi)) + p_i(0) + p_b(U_b - .01(\pi)) \tag{29}
\]

\[
E[\text{Payoff}(X^S)] = p_g(U_g + .051\bar{\delta}(\pi)) + p_i(0) + p_b(U_b - .051\bar{\delta}(\pi)) \tag{30}
\]

Kurt can determine how many others he must interact with to reach his desired probability of promotion (influenced by his native ability) and with whom exactly he will interact. Consider the case that Kurt is a nervous fellow and can only rest happily once he has a 100\% chance of promotion, and that he has a native ability of .5. Kurt then needs quantities of regular and supervisory employees in his social network such that \(.01(R) + .051\bar{\delta}(S) \geq .5\). Kurt will
interact with the employee from whom he expects the highest payoff, then the second highest, until his social network has reached the size to make the inequality hold.

In this gender-homogenous firm, if the probabilities of a good interaction with individuals are randomly distributed across possible pairings, the model then predicts that each person will have the same estimated number of interactions. If the probabilities are not random but affected by how good an employee’s personality is, then the model predicts that genial employees will be more likely to interact successfully and be chosen for interactions. Regardless of which condition holds – common sense suggests elements of both in reality – the model can reasonably depict individuals interacting heterogeneously.

4.2 A two-gender workplace with potential for harassment – the harasser’s decision

We now reintroduce gender to the model and consider a workplace that employs both genders and in which the possibility of harassment exists. Adjusting Equation 6 from Section 3.3, the payoff of initiating a harassing interaction for a Type H man who is a regular employee with a female regular employee is now as follows:

\[
E[Payoff_1(X)] = (p_g - h_1) \left( U_g + .01(\pi) \right) + p_i(0) + (p_b + h_1)(U_b - .01(\pi)) + Satisfactions + a \ast (h_1) \ast (Consequences) \tag{31}
\]

\(a \ast h_1\) is the probability that the female will report this interaction as harassment.

\(0 \leq a \leq 1\)

\(\pi\) is the value of the promotion

All variables are as described previously in this section and Section 3.3

In contrast, for a Type G or Type H man initiating a normal interaction with a female regular employee, the payoff is:
\[ E[\text{Payoff}_2(X)] = (p_g - h_2) \left( U_g + .01(\pi) \right) + p_l(0) + (p_b + h_2)(U_b - .01(\pi)) + a \cdot h_2(\text{Consequences}) \]  \hspace{1cm} (32)

For a Type H man deciding between the two, the difference is:

\[-(h_1 - h_2) \left( U_g + .01(\pi) \right) + (h_1 - h_2)(U_b - .01(\pi)) + \text{Satisfactions} + a(h_1 - h_2)(\text{Consequences}) \]  \hspace{1cm} (33)

For him to decide to harass, it must be the case that:

\[ \text{Satisfactions} \geq (h_1 - h_2) \cdot [(U_b - .01\pi) - (U_g + .01\pi) + a(\text{Consequences})] \]  \hspace{1cm} (34)

The Type H man must receive more intrinsic benefit from harassing than the change in the probabilities of outcome due to his harassing behavior.

4.3 A two-gender workplace with potential for harassment – the non-harasser’s decision

As discussed earlier in this section, a regular employee decides with whom to interact based on the expected payoff of the interaction, interacting with each person until he has reached the desired probability of promotion. When the possibility of harassment exists with some of these employees, the lower expected payoff will always shift a woman lower in the ranking than she would have been, meaning that any given man will choose to interact with other men over a woman because of the potential for false accusation of harassment. The difference between the expected payoffs of an interaction with a man, compared to an otherwise identical woman, is:

\[ h_2 \left[ (U_g + .01\pi) - (U_b - .01\pi) + a \cdot \text{Consequences} \right] \]  \hspace{1cm} (35)

This result is equal to the probability of being mistaken as harassing times the difference between the good and bad outcomes, including the effect on the probability of receiving a promotion and possible consequences if the interaction is reported.
Using the same workplace structure (100 employees total, 24 of which are supervisors), consider a specific numerical example in the case that there is only one female regular employee and no female supervisors. Further, assume each regular employee has a level of native ability and a desired probability of promotion (see Equation 21) such that s/he must interact with at least 10 regular employees to reach that probability. Assume that the sole female, Lisa, had the best personality of all the employees in the office and would be ranked first to interact with for a certain co-worker, Leopold. Let us assign the following numerical values to the interaction equation between Lisa and Leopold:

<table>
<thead>
<tr>
<th>Variable (interpretation)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi$ (net present value of promotion)</td>
<td>$50,000</td>
</tr>
<tr>
<td>$p_g$ (probability of good interaction)</td>
<td>.75</td>
</tr>
<tr>
<td>$p_i$ (probability of indifferent interaction)</td>
<td>.1</td>
</tr>
<tr>
<td>$p_b$ (probability of bad interaction)</td>
<td>.15</td>
</tr>
<tr>
<td>$U_g$ (non-income benefit of good interaction)</td>
<td>0</td>
</tr>
<tr>
<td>$U_b$ (non-income harm of bad interaction)</td>
<td>0</td>
</tr>
<tr>
<td>Consequences (cost of having harassment reported against you)</td>
<td>$50,000</td>
</tr>
<tr>
<td>$a$ (probability that perceived harassment is reported)</td>
<td>.5</td>
</tr>
<tr>
<td>$h_1$ (probability harassment is perceived as such)</td>
<td>.75</td>
</tr>
<tr>
<td>$h_2$ (probability that non-harassment is perceived as harassment)</td>
<td>.01</td>
</tr>
</tbody>
</table>

In the absence of any concerns about sexual harassment (if Lisa were a man, for example), Leopold’s expected payoff from interacting with her would be as follows (see Equation 24 for the general form):
Because the payoff for this interaction is both positive and large, Leopold would have a strong incentive to interact with Lisa. Instead, because of the risk of a ‘false positive’ perception of harassment, Leopold’s expected payoff is:

\[
E(X^R) = (.75 - .01)(.01 \times 30,000) + (.1 + .01)(-.01 \times 30,000) + .5(.01 \times -50,000) \\
= -$61
\]

This example vividly illustrates that even if there is a small possibility that an innocent interaction is mistaken as harassment (and an even smaller possibility it is reported as such), the presence of the risk can drastically affect the expected value of an interaction. Leopold no longer has any positive incentive to interact with Lisa and has a good incentive to avoid interaction with her because of the penalty if she considers it harassment and the decreased probability of receiving the promotion.

Since Lisa is the only female in this workplace, most male employees could interact with others instead of her; instead of being the person most interacted with, she could be the person the least interacted with (assuming there are no genuinely unpleasant male employees) and thus be unable to expand her social network to increase her probability of promotion. While Leopold could ignore Lisa to talk to Larry or Lester, Lisa will become a workplace pariah. While the example is extreme in reducing the number of female employees to one, the example extrapolates to a workplace or subgroup therein in which there are few women relative to the number of men. This result is particularly troubling if workplaces with few females are more likely to have problems with harassment and thus to impose harsh penalties for it. Further, this suggests that anti-harassment policies may impede the advancement of women in male-
concentrated professions, when it may be a social goal to increase the number of women
promoted in such professions.

The network effects of skewing the composition of a workplace will be examined further
in Section 4.5, but first we will examine the decision made by the other kind of employee in the
model – the supervisor.

4.4 A two-gender workplace with potential for harassment – supervisor-employee
interactions

In this model, an employee gains more from adding a supervisor to his/her social network
than by adding a regular employee, an intuitive assumption because of the supervisor’s influence
on whether or not the employee is promoted. When a supervisor decides whether to interact with
a regular employee, there is no career benefit in terms of promotion, as this workplace has only
two tiers, but there is a benefit to having acquired information about the regular employee. This
information will enable the supervisor to better recommend/decide whether or not this employee
should be promoted, to evaluate the employee, or to make enforcement less costly. The
supervisor’s communication decision, therefore, contains a fixed benefit from such interactions.
A supervisor’s decision to interact with a regular employee of the same gender would have the
following expected payoff, similar to the payoff demonstrated in Equation 1:

\[ E[Payoff(X^R)] = V + p_g(X_g) + p_i(0) + p_b(X_b) \text{ s.t. } p_g + p_i + p_b = 1 \]  

(36)

\( X \) is the social interaction with the regular employee
\( V \) is the value of the information acquired as a result of the interactions
The \( X_g \)s and \( p_g \)s are as in previous equations

The different benefits of good and bad interactions are, unlike the previous section and like the
original model of Section 3, simply based on the supervisor’s intrinsic valuation of the social
outcome. The element of “business” exists solely in $V$, the value to the supervisor of learning about the employee through interaction. This payoff equation also holds for a female supervisor interacting with a male regular employee – for now, we assume she need not fear harassment because presumably, no man will harass his supervisor because she could immediately fire him. (The model makes no allowances for arrant stupidity, which may be a highly unrealistic omission.)

A male supervisor’s expected payoff with a female regular employee for a non-harassing interaction is:

$$E[Payoff(X)] = V + (p_g - h_2)(X_g) + (p_b + h_2)(X_b) + a * h_2(Consequences + QPQ) \tag{37}$$

The additional term attached to $Consequences$ recognizes that a supervisor faces greater consequences for harassment both because of his station and because a court could perceive any harassment from a supervisor as “quid pro quo” harassment, a possibility, however small, that does not exist for potential harassment between colleagues. (Of course, the supervisor may have a higher value of $Satisfactions$ precisely because of this status difference.) If a male supervisor is Type H and thus has a taste for harassment, his expected payoff for a harassing interaction is:

$$E[Payoff_{f_1}(X)] =$$

$$V + (p_g - h_1)(X_g) + (p_b + h_1)(X_b) + Satisfactions + a * h_1(Consequences + QPQ) \tag{38}$$

The results are similar to those of Part 3 – a male supervisor has less of an incentive to interact with a female regular employee than with a male regular employee who is otherwise identical, unless the supervisor has a taste for harassment and the benefit he receives from harassment ($Satisfactions$) outweighs the increased risk of consequences ($h_1 > h_2$). Such a result has been
observed in practice. An article in The New York Times examined why law firms have few female partners relative to the number of female associates and commented,

Some of this give-and-take enters gray areas that may have as much to do with caution as it does with biases. Is a male boss reluctant to invite a younger female lawyer out for a drink because water cooler chatter might spark rumors of an affair or give rise to a sexual harassment suit? (O’Brien 2006).

This description supports the existence of a fear of mistaken harassment and the negative consequences that result for female employees.

For a male regular employee interacting with male supervisor or a female regular employee interacting with female supervisor, the expected payoff is Equation 30 as described in Section 4.1:

\[
E[Payoff(X^S)] = p_g (U_g + 0.051\bar{\delta}(\pi)) + p_i(0) + p_b(U_b - 0.051\bar{\delta}(\pi))
\]

A female regular employee interacting with a male supervisor risks harassment:

\[
E[Payoff(X^S)] = (p_g - h)(U_g + 0.051\bar{\delta}(\pi)) + p_i(0) + (p_b + h)(U_b - 0.051\bar{\delta}(\pi)) + h(Harassment)
\]

\(h\) is the probability with which she expects to experience harassment. \(Harassment\) is the negative experience the woman undergoes from undergoing what she perceives as harassment; \(Harassment \leq 0\).

The female regular employee faces a greater cost to interacting with a male supervisor than with a female supervisor – a cost that the male regular employee does not face with either.

This section has not yet considered the scenario in which a subordinate is required to interact with a supervisor because of the job. Jean Tirole describes a model of hierarchy and bureaucracy in which employees exogenously choose a level of effort and managers both
observe and report employee effort.\textsuperscript{13} Tirole simplifies to the case that managers either fully observe employee productivity or do not observe it at all. If we assume that the managers are the principals of the firm (as is the case in law firms, for example), the model can be further simplified to ignore concerns of reporting.

Tirole’s ideas fit within the context of this model if we consider the case that a manager observes productivity based on interactions of the kind described so far. We can thus re-imagine Equation 36 so that the variable $V$, the value of the information acquired from the interaction, includes value of learning the employee’s productivity level. In addition, the employee’s decision to interact with a supervisor can be modified to include the benefit of keeping his/her job or the benefit of making his/her productivity level known. In this case, if both fear of revealing a low productivity level and fear of harassment lead to an employee interacting less with a supervisor, the employees who fear either harassment from supervisors will be interpreted to have lower productivity levels.

Such a result supports the familiar conclusion that harassment harms a woman’s career and productivity. In addition, the effect could build upon itself: if harassment leads a female to a decreased probability that her productivity will be observed, she will have a decreased incentive to be productive. When she is observed, perhaps when interacting with a female manager, her observed productivity will be lower than it would have been otherwise, further hurting the female employee.

\textsuperscript{13} For more discussion, see Tirole (1986).
However, even under these modified assumptions, anti-harassment policy may again have the same negative career effects for females as harassment itself. If a supervisor interacts less with female employees because he fears his interaction will be mistaken as harassment, he is less likely to know their true productivity levels and the same effect could occur for the opposite reason. Female employees would again be harmed.

4.5 Network effects

Previously, we discussed the probability a woman would interpret an interaction as harassment \((h_1 \text{ or } h_2)\) without determining the value of these parameters within the context of the model. Now, let us consider the case that these \(h\)s are partially determined by the total number of harassers in the workplace in addition to whether the individual himself is a harasser or not:

\[
h_1 = \frac{\# \text{ of } H}{\# \text{ of male employees}} + c \tag{40}
\]

\(h_1\) remains the probability that a female will interpret the harassing interaction as such. 
\(\# \text{ of } H\) is the number of Type H men working at the firm.
\(c\) is some constant between 0 and \(1 - \left(\frac{\# \text{ of } H}{\# \text{ of males}}\right)\) that represents the increased likelihood that an actually harassing interaction will be perceived as one compared to an innocent interaction.

Similarly, the probability that a non-harassing interaction will be interpreted as such is:

\[
h_2 = \frac{\# \text{ of } H}{\# \text{ of male employees}} \tag{41}
\]

Modifying Equation 31 from Section 4.2, a Type H man who is a regular employee receives the following payoff from initiating a harassing interaction with a female regular employee:
is the probability that the female will report this interaction as harassment.
0 \leq a \leq 1
\pi is the value of the promotion
All variables are as described previously in this section and Section 3.3

In contrast, the payoff for either a Type G or Type H man initiating a normal interaction with a female regular employee is:

\[
E[Payoff_{f_2}(X^R)] = (p_g - \left( \frac{\# \text{ of } H}{\# \text{ of male employees}} + c \right)) (U_g + .01(\pi)) + p_i(0) + (p_b + \left( \frac{\# \text{ of } H}{\# \text{ of male employees}} + c \right)) (U_b - .01(\pi)) + \text{Satisfactions} + a \star \left( \frac{\# \text{ of } H}{\# \text{ of male employees}} + c \right) \star \text{(Consequences)}
\] (43)

For a Type H man deciding between the two, the difference is now measured as:

\[
-c \left( U_g + .01(\pi) \right) + c \left( U_b - .01(\pi) \right) + \text{Satisfactions} + ac(\text{Consequences})
\] (44)

For the Type H man to decide to harass, it must be the case that:

\[
\text{Satisfactions} \geq c \star \left[ (U_b - .01\pi) - (U_g + .01\pi) + a(\text{Consequences}) \right]
\] (45)

We can now derive the change in a man’s expected payoff to a harassing interaction with respect to the number of Type H men in a firm:

\[
\frac{\delta(E(X^R))}{\delta \left( \frac{\# \text{ of } H}{\# \text{ of male employees}} \right)} = -(U_g + .01\pi) + (U_b - .01\pi) + a \star \text{Consequences}
\]

\[
= (U_b - U_g) - .02\pi + a \star \text{Consequences}
\] (46)
The term \((U_b - U_g)\), if nonzero, would be a negative number under standard assumptions of the parameters. In addition, both the terms \(-0.02\pi\) and \(a \ast Consequences\) would be negative: the first because people positively value the promotion and the second because they negatively value the consequences of reported harassment. Thus, as the number of Type H men in a firm increases, the relative incentive to harassment actually decreases. Because \(h_1\) is determined by the number of Type H men in the firm, if the firm had but one Type H man, he would harass happily because the negative effects (to himself) of his harassment would be much lower, and the enjoyment he received from harassing would be the same.

The change in the expected payoff of a non-harassing interaction with respect to the percent of Type H men is the same expression: \((U_b - U_g) - 0.02\pi + a \ast Consequences\). Notably, if the number of Type H men is zero, \(h_2 = 0\) and there is no possibility that an innocent interaction will be mistaken as harassment. The presence of a Type H man in the firm reduces the payoff for any interaction between a man and a woman; the proverbial bad apple does, in a certain sense, spoil the whole bunch. As the number of Type H men increases, the probability of mistake increases. It seems reasonable to imagine that if a firm had many harassers, a female employee would be more likely to expect and suspect harassment.

This result of the model demonstrates the benefits of reducing the number of Type H men present in the workplace. In practice, this suggests that harassment training could benefit a workplace if such training could remove a man’s preference for harassment, reducing \(Satisfactions\) to 0 and changing him from a Type H to a Type G man. Whether this is possible is a psychological rather than economic question – if it is, the findings of Antecol and Cobb-Clark
(2003) that harassment training increases a man’s tendency to perceive an interaction as harassment could indicate that training effected such a change. Regardless, the model demonstrates that eliminating Type H men from the workplace prevents harassment while still encouraging innocent communication more effectively than can be accomplished by simply punishing harassers. This suggests that a screening mechanism of some kind would be an effective policy under the model’s set of assumptions so far.

The change in the expected payoff of a harassing interaction with respect to the value of the promotion ($\pi$) is as follows:

$$\frac{\delta(E(x^H))}{\delta(\pi)} = .01 \left( p_g - p_b - 2 \left[ \frac{\# of H}{\# of male employees} \right] - 2c \right)$$

Here, .01 is the amount by which a positive interaction with another regular employee increases the probability of promotion. The magnitude of the promotions’ influence on the interaction’s expected payoff depends on the probabilities of the various forms of interaction: if the probability of a good interaction is high, the value of the promotion greatly affects the expected payoff of the interaction because the “cost” of the interaction, the risk of being reported, is unaffected by $\pi$. If the probability of a bad interaction is higher, the value of the promotion will matter less because of the quasi-fixed risk of reporting. If the probability of a bad interaction is very high, a larger value of the promotion would negatively affect the expected payoff because the risk of hurting one’s chances of the promotion would be too great for someone to attempt the interaction. A valuable promotion can deter harassment by increasing its cost to the harasser.

In comparison, the change in the expected payoff of a non-harassing interaction is:

$$\frac{\delta(E(x^N))}{\delta(\pi)} = .01 \left( p_g - p_b - 2 \left[ \frac{\# of H}{\# of male employees} \right] \right)$$
The difference between the two derivatives lies in the term $2c$, as $c$ is the difference between a true positive and a false positive interpretation of harassment. This is an extremely intuitive result – the difference in perceptions between a harassing interaction and a non-harassing one results in the difference in the change in the cost to the harassing interaction compared to the non-harassing one. This result suggests that if companies attempt to combat harassment using an incentive similar to an efficiency wage, the policy’s success in deterring harassment without deterring non-harassing communication will depend on how differently the two kinds of interaction are perceived. If there is little distinction between the two, a valuable promotion serves to deter both kinds of communication and works in the same way as a large punishment for harassment would – in these cases, men interact less with women if doing so could cost them.

The similarity of the cases of large promotions and large punishments can be demonstrated by the derivatives of the expected payoffs of the two kinds of interactions with respect to $\text{Consequences}$. For a harassing interaction:

$$\frac{\delta(E(X^H_1))}{\delta(\text{Consequences})} = a\left(\frac{\# \text{ of } H}{\# \text{ of male employees}} + c\right) \quad (49)$$

The effect of a change in the consequences of being reported for harassment is determined by the probabilities that harassment will be perceived as such and that it will be reported. The derivative of a non-harassing interaction is similarly determined by these probabilities:

$$\frac{\delta(E(X^H_2))}{\delta(\text{Consequences})} = a\left(\frac{\# \text{ of } H}{\# \text{ of male employees}}\right) \quad (50)$$

The difference, again, exists in the term $c$, the difference between a true positive and a false positive perception of harassment. If a company chooses a policy of increasing the
consequences of harassment, such an increase will create a greater disincentive for harassment than for regular interaction only by the amount of the distinction between the two.

5. CONCLUSIONS

Anti-harassment policy may lead to serious unintended consequences because of the possibility of mistake. This model formalizes the delicate trade-off between deterring harassment and obstructing beneficial communication. The benefits of communication are reduced both if a potential victim fears harassment and if a potential accused party fears mistake of harassment. Penalties for harassment can stop innocent communication particularly well when harassment and non-harassment are easily confused for each other, a problem unique to harassment among most other torts. It is easier to show that someone broke your leg than to show that they demeaned your person. In the extreme case that whether an interaction is called harassment is unrelated to intention to harass, a Nash equilibrium occurs where men choose always to harass and women choose never to report it.

This theoretical result leads to a conclusion similar to that of Antecol and Cobb-Clark’s (2006) result: that whether a person feels harassed is more important than whether or not she has is harassed based on some empirical standard. While policies that change perceptions of potential harassers may be helpful, policies increasing the likelihood that a potential victim feels harassed are counterproductive. Companies should avoid encouraging women to hypersensitivity lest they encourage her to mistake non-harassing interactions for harassing ones and increase the probability that they feels harassed, decreasing their overall happiness.
In the case that benefits from communication are linked to an individual’s probability of promotion, a vigorous anti-harassment regime can lead to the ostracization of women in the workplace because of the risk inherent in a male-female interaction. As women become the least desirable colleagues for an employee to add to his social network, women may become less likely to be promoted than equally- or less-qualified men. Even small probabilities of mistake can drastically change incentives to interact. The structure of a workplace social network effects these results in other ways, as well: when this probability is determined by the number of male employees with a taste for harassment, increasing the relative percentage of harassers reduces incentives for employees to interact in both a harassing and non-harassing way.

These results have implications for ways to structure policies targeted at countering sexual harassment. Following this model, companies would do better to focus training on genuinely differentiating harassment from non-harassment rather than simply emphasizing as wrong anything that might be perceived as harassment. In addition, companies should not make the penalty for harassment too severe, nor give undue credence to every harassment claim, lest they deter innocent communication at the expense of the employees the policies are designed to help.

The model does not apply exclusively to sexual harassment but could extrapolate to any situation involving misbehavior, reporting, and the potential for mistake. One such example is police misconduct and the ways in which it is reported. Medical malpractice also might fit this model. Though these examples lack some of the subjectivity particular to harassment, they similarly feature cases in which the reporting is linked to the social relationships among
colleagues. In these situations, uncertainty and heavy consequences combine to cause problems and unintended consequences.

Future research could focus on empirical tests of this model both in psychology and in economics; for a problem such as this, some intersection between the two, such as an economic experiment, could provide data on individual decisions not available otherwise. Further extension of the topic could contribute to examination of the “glass ceiling” in the workforce; fear of harassment accusations could contribute to the fact that women make up a significant part of the labor force but still hold disproportionately fewer upper-level positions in many industries and professions. Another possible study could regress various kinds of anti-harassment training on number of harassment complaints filed and number of complaints dismissed.

The dilemma of deterring harassment without deterring communication may not be wholly intractable in the long run, even if current federal law remains unchanged. Firms either face the cost of harassment lawsuits and fines, or else the cost of lost employee interaction, motivating them to eliminate both such costs. If the possibility of mistaken harassment is indeed determined by the number of employees with a taste for harassing, employers will have an incentive to identify harassers before they are hired and to exclude them from the workplace entirely. This would in turn increase the cost to keeping one’s taste for harassing, eventually (if one is optimistic) leading to a world in which the fear of harassment becomes unnecessary. If so, this result would be an economist’s favorite happy ending: whether one considers policy to have spurred or impeded the change, the market may, at last, correct for this problematic, productivity-reducing characteristic.
APPENDIX 1 – GLOSSARY OF SYMBOLS

Type G: A man who does not benefit from or enjoy sexually harassing others

Type H: A man who receives some benefit or enjoyment from sexually harassing others

$X$: A social interaction

$X_g$: The value of a social interaction with a good or positive result

$X_b$: The value of a social interaction with a bad or negative result

$p_g$: The probability of a successful interaction (which results in $X_g$)

$p_i$: The probability that the partner is indifferent to the interaction (no benefit or harm)

$p_b$: The probability of a bad interaction (which results in $X_b$)

$h$: The probability that the female will experience what she interprets as harassment during an interaction in the case that she cannot distinguish between Type G and Type H men; $0 \leq h \leq 1$

$Harassment$: The disutility a woman undergoes from experiencing what she perceives as harassment; $Harassment \leq 0$.

$h_a$: The probability that the female will experience harassment during her interaction with a Type H male in the case that she can distinguish between types of men.

$h_b$: is the probability that the female will experience harassment during her interaction with the Type G male in the case that she can distinguish between types of men; $h_b < h_a$

$h_f$: The probability that the female recipient of the interaction will interpret a harassing interaction as harassment.

$a$: The probability that a female who considers an interaction to be harassment will report the interaction as such; $0 \leq a \leq 1$

$Satisfactions$: The benefit that a Type H man receives from the harassing interaction (for a Type G man, $Satisfactions \leq 0$)

$Consequences$: The negative ramifications such as investigation, litigation, losing a job, etc. that (may) result from being reported for harassment.

$h_2$: The probability that the recipient of an interaction not intended as harassment will nonetheless interpret the interaction as harassment; $h_2 < h_f$

$P_1$: The probability that Employee 1 receives a promotion
\(T\): The total number of employees in an employee’s social network

\(S\): The number of supervisors in an employee’s social network

ability: An employee’s native ability

\(X^R\): A regular employee’s interaction with another regular employee

\(X^S\): A regular employee’s interaction with a supervisor

\(\pi\): The value of a promotion to an employee

\(U_g\): The non-promotion-related happiness that Kurt receives from a good interaction

\(U_b\): The non-promotion-related unhappiness that Kurt receives from a bad interaction

\(V\): The value of the information that a supervisor acquires by interacting with a regular employee

\(QPQ\): The additional consequences that a supervisor receives when reported as harassing a regular employee
APPENDIX 2 – LIST OF EQUATIONS AND FIRST-ORDER CONDITIONS

Section 3

3.1 Model of communication between employees

1. \[ E[Payoff_1(X)] = p_g(X_g) + p_i(0) + p_b(X_b) \]

3.2 Communication and the possibility of harassment: female-initiated interactions

2. \[ E[Payoff_1(X)] = (p_g - h)(X_g) + p_i(0) + (p_b + h)(X_b) + hHarassment \]
3. \[ E[Payoff_1(X^H)] = (p_g - h_a)(X_g) + p_i(0) + (p_b + h_1)(X_b) + hHarassment \]
4. \[ E[Payoff_1(X^G)] = (p_g - h_b)(X_g) + p_i(0) + (p_b + h_2)(X_b) + hHarassment \]
5. \[ \frac{dE}{dh} = -X_g + X_b + Harassment \]

3.3 Communication and the possibility of harassment: male-initiated interactions

6. \[ E[Payoff_1(X)] = (p_g - h_1)(X_g) + p_i(0) + (p_b + h_1)(X_b) + Satisfactions + a * h_1(Consequences) \]
7. \[ E[Payoff_1(X)] = (p_g - h_1)(X_g) + p_i(0) + (p_b + h_1)(X_b) + a * h_1(Consequences) \]
8. \[ E[Payoff(X)] = (p_g - h_2)(X_g) + p_i(0) + (p_b + h_2)(X_b) + a * h_2(Consequences) \]
9. \[ (h_1 - h_2)(X_g + X_b + a * Consequences) \]
10. \[ Satisfactions = (h_1 - h_2)(X_g - X_b - aConsequences) \]
11. \[ f = (h_1 - h_2)(X_g - X_b - aConsequences) - Satisfactions \]
12. \[ \frac{df}{dh_1} = X_g - X_b - aConsequences \]
13. \[ \frac{df}{dConsequences} = (h_2 - h_1)a \]
14. \[ \frac{df}{da} = (h_2 - h_1)Consequences \]
15. \[ \frac{df}{dh_2} = -X_g + X_b + a * Consequences \]

3.4 Communication and the possibility of harassment: interactions initiated by Type G males

16. \[ \frac{dE}{dh_2} = -X_g + X_b + a * Consequences \]
17. \[ \frac{dE}{da} = h_2(Consequences) \]
18. \[ \frac{dE}{dConsequences} = (h_2)a \]
Section 4

19. \[ P_1 = \frac{T}{\# \text{of employees}} + \frac{S}{\# \text{of supervisors}} + \text{ability} \]
20. \[ P_1 = \frac{T}{100} + \frac{S}{24} + \text{ability} \]
21. \[ P_1 = .01(R) + .0516(S) + \text{ability} \]
22. \[ \frac{\partial P}{\partial R} = .01 \]
23. \[ \frac{\partial P}{\partial S} = .0516 \]

4.1 A single-gender workplace

24. \[ E[Payoff_f_1(X)] = p_g(X_g) + p_i(0) + p_b(X_b) \]
25. \[ X_g^R = U_g + .01(\pi) \]
26. \[ X_b^R = U_b - .01(\pi) \]
27. \[ X_g^S = U_g + .0516(\pi) \]
28. \[ X_b^S = U_b - .0516(\pi) \]
29. \[ E[Payoff(X^R)] = p_g(U_g + .01(\pi)) + p_i(0) + p_b(U_b - .01(\pi)) \]
30. \[ E[Payoff(X^S)] = p_g(U_g + .0516(\pi)) + p_i(0) + p_b(U_b - .0516(\pi)) \]

4.2 A two-gender workplace with potential for harassment – the harasser’s decision

31. \[ E[Payoff_f_1(X)] = (p_g - h_1)(U_g + .01(\pi)) + p_i(0) + (p_b + h_1)(U_b - .01(\pi)) + \text{Satisfactions} + a * (h_1) * (\text{Consequences}) \]
32. \[ E[Payoff_f_2(X)] = (p_g - h_2)(U_g + .01(\pi)) + p_i(0) + (p_b + h_2)(U_b - .01(\pi)) + a * h_2(\text{Consequences}) \]
33. \[ -(h_1 - h_2)(U_g + .01(\pi)) + (h_1 - h_2)(U_b - .01(\pi)) + \text{Satisfactions} + a(h_1 - h_2)(\text{Consequences}) \]
34. \[ \text{Satisfactions} \geq (h_1 - h_2) * [(U_b - .01\pi) - (U_g + .01\pi) + a(\text{Consequences})] \]

4.3 A two-gender workplace with potential for harassment – the non-harasser’s decision

35. \[ h_2[(U_g + .01\pi) - (U_b - .01\pi) + a \, \text{Consequences}] \]

4.4 A two-gender workplace with potential for harassment – supervisor-employee interactions

36. \[ E[Payoff(X^R)] = V + p_g(X_g) + p_i(0) + p_b(X_b) \]
37. \[ E[Payoff(X)] = V + (p_g - h_2)(X_g) + (p_b + h_2)(X_b) + a * h_2(\text{Consequences} + \text{QPQ}) \]
38. \[ E[\text{Payoff}_1(X)] = V + (p_g - h_1)(X_g) + (p_b + h_1)(X_b) + \text{Satisfactions} + a \times h_1(\text{Consequences} + \text{QPQ}) \]

39. \[ E[\text{Payoff}_2(X^S)] = (p_g - h)(U_g + .051\bar{\delta}(\pi)) + p_l(0) + (p_b + h)(U_b - .051\bar{\delta}(\pi)) + h(\text{Harassment}) \]

4.5 Network effects

40. \[ h_1 = \frac{\# \text{ of } H}{\# \text{ of male employees}} + c \]

41. \[ h_2 = \frac{\# \text{ of female employees}}{\# \text{ of H}} \]

42. \[ E[\text{Payoff}_1(X)] = (p_g - \left(\frac{\# \text{ of } H}{\# \text{ of male employees}} + c\right))(U_g + .01(\pi)) + p_l(0) + (p_b + \left(\frac{\# \text{ of } H}{\# \text{ of male employees}} + c\right))(U_b - .01(\pi)) + \text{Satisfactions} + a \times \left(\frac{\# \text{ of } H}{\# \text{ of male employees}} + c \times (\text{Consequences}) \right) \]

43. \[ E[\text{Payoff}_2(X^R)] = (p_g - \left(\frac{\# \text{ of } H}{\# \text{ of male employees}}\right))(U_g + .01(\pi)) + p_l(0) + (p_b + \left(\frac{\# \text{ of } H}{\# \text{ of male employees}}\right))(U_b - .01(\pi)) + a \times \left(\frac{\# \text{ of } H}{\# \text{ of male employees}} \right) \times (\text{Consequences}) \]

44. \[ -c(U_g + .01(\pi)) + c(U_b - .01(\pi)) + \text{Satisfactions} + ac(\text{Consequences}) \]

45. \[ \text{Satisfactions} \geq c \times [(U_b - .01\pi) - (U_g + .01\pi) + a(\text{Consequences})] \]

46. \[ \frac{\delta(E(X^R))}{\delta(\pi)} = (U_b - U_g) - .02\pi + a \times \text{Consequences} \]

47. \[ \frac{\delta(E(X^R))}{\delta(\pi)} = .01p_g - p_b - 2\left[\frac{\# \text{ of } H}{\# \text{ of male employees}}\right] - 2c \]

48. \[ \frac{\delta(E(X^R))}{\delta(\pi)} = .01(p_g - p_b - 2\left[\frac{\# \text{ of } H}{\# \text{ of male employees}}\right]) \]

49. \[ \frac{\delta(\text{Consequences})}{\delta(E(X^R))} = a \left(\frac{\# \text{ of } H}{\# \text{ of male employees}} + c\right) \]

50. \[ \frac{\delta(\text{Consequences})}{\delta(E(X^R))} = a \left(\frac{\# \text{ of } H}{\# \text{ of male employees}}\right) \]
REFERENCES


