# A Theory of Civil Disobedience

## Faculty Research Working Paper Series

Edward L. Glaeser

Harvard Kennedy School

Cass R. Sunstein

Harvard Law School

**July 2015**

**RWP15-036**

## Acknowledgements

Glaeser thanks the Taubman Center for State and Local Government for financial support. Seminar participants in the University of Michigan Law School provided helpful comments. Yueran Ma provided superb research assistance.

The views expressed herein are those of the authors and do not necessarily reflect the views of the Harvard Kennedy School or the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w21338.ack



## Abstract

From the streets of Hong Kong to Ferguson, Missouri, civil disobedience has again become newsworthy. What explains the prevalence and extremity of acts of civil disobedience? This paper presents a model in which protest planners choose the nature of the disturbance hoping to influence voters (or other decision-makers in less democratic regimes) both through the size of the unrest and by generating a response. The model suggests that protesters will either choose a mild "epsilon" protest, such as a peaceful march, which serves mainly to signal the size of the disgruntled population, or a "sweet spot" protest, which is painful enough to generate a response but not painful enough so that an aggressive response is universally applauded. Since non-epsilon protests serve primarily to signal the leaders' type, they will occur either when protesters have private information about the leader's type or when the distribution of voters' preferences are convex in a way that leads the revelation of uncertainty to increase the probability of regime change. The requirements needed for rational civil disobedience seem not to hold in many world settings, and so we explore ways in which bounded rationality by protesters, voters, and incumbent leaders can also explain civil disobedience.

## I. Introduction

In both democratic and undemocratic nations, political protestors sometimes engage in civil disobedience. They break the law in order to register their protest, often with the hope of increasing the likelihood of significant reform. A particularly interesting feature of civil disobedience is a distinctive motivation, which is *to produce an aggressive response from the relevant authority, which will in turn lead to a heightened sense, on the part of the public as a whole, that the authority needs to be reformed or replaced*. On this view, the goal of civil disobedience is both to deepen and to widen public concern through the adverse reaction that it provokes, and through that route to produce large-scale reform.

In general, achievement of that goal would seem to be most unlikely. Those who break the law usually fail to win public support. On the contrary, they discredit themselves; they produce widespread opprobrium. When they are met with force, the public approves. But some cases, law-breakers succeed. Protestors rarely invoke Lenin's idea of "heightening the contradictions," but in some cases, they seem to do exactly that, and ultimately have significant effects. Why and when? Our purpose in this Article is to answer that question.

As we understand it here, civil disobedience is about signaling, and in two different ways. *First*, the disobedients (as we shall call those protestors who break the law) signal that they are displeased with the governing authority, and in particular that the authority is responsible for serious mistakes and injustice. That signal might have a major influence on other citizens, which is to alter their own judgments (by informing them of what other people think) and also to reduce pluralistic ignorance (people's ignorance about the beliefs and preferences of other people). As in the conventional models of protest (Granovetter 1978; Lohmann 1994, DiPasquale and Glaeser, 1998), so too for the disobedients: They might create a kind of cascade in which large numbers of people ultimately "tip" (cf. Kuran, 1998; Murray, 2015). At the same time (and we see this as particularly important), disobedience can make injustice salient when it might otherwise be seen as some kind of (inevitable) background fact. In these respects, disobedience should be seen as a strong form of ordinary protest activity, in which the "volume" of the action is increased (and potentially greatly so) because it is inconsistent with law. If citizens see that some or many others are willing to risk sanctions, they will have an enhanced sense of the intensity of current disapproval of the status quo, thus altering informational signals (about its true character) and also reputational incentives (by revealing the views of fellow citizens). By itself, this signal might have a sufficient effect on the authority, who might respond by moving in the desired direction.

*Second* (and this is a central part of our focus here), the disobedients sometimes seek to provoke the authority to signal its own bad character or type. Of course it is usually best, from the standpoint of the disobedients, if the authority changes the status quo in the desired way (in the extreme case, by relinquishing authority). But if that is not possible, the disobedients might

want instead to provoke, from the authority, the revealing signal.[1] To achieve that end, the disobedience must be damaging enough to elicit a forceful response, providing that signal, but if it is too damaging, there is a grave risk, which is that forceful responses will seem fully justifiable and therefore welcome. "Damage" can be understood both in terms of the particular law that is being violated and the number of people who are violating it. If, for example, the disobedient commit murder, rape, or assault, citizens will ordinarily welcome a forceful response[2]; if the disobedient walk the streets at a time when they are not permitted to do so, the use of force will be less likely to be well-received. And if 500 people walk the streets, force will be far less welcome than if the streets are blanketed by 100,000 (which may bring productive activity to a halt).

The disobedients must therefore find some kind of "sweet spot" in which their action is sufficient to provoke either widespread sympathy or a forceful response while ensuring that that response contains the desired signal about the character of the authority. The authority must find sweet spots of its own, ignoring certain disobedience (on the ground that a reaction would fuel the relevant movement) but responding sufficiently to other disobedience (on the ground that passivity would allow dangerous growth). As we shall see, these conclusions have strong implications not only for the responses of rational authorities but also for enforcement activity and prosecutorial discretion in the face of civil disobedience (see Dworkin, 1967). Prosecutions can themselves create a desirable signal for the disobedient, and prosecutors should be aware of that risk in thinking about the best way to deter conduct.

Much of our analysis will elaborate on the two signals associated with civil disobedience, with the assumption that the relevant actors are rational. In many cases, the assumption tracks reality, or at least it is close enough. The disobedients, or their leaders, can be highly strategic actors; Martin Luther King, Jr. is a prime example. In other cases, however, psychological or behavioral factors much complicate the analysis. For example, the disobedients might well be outraged, and their outrage might compromise their strategic goals; expression of outrage might seem an end in itself, even if it is unlikely to produce good consequences. (In fact, an apparently noninstrument motivation for engaging in civil disobedience might be helpful in instrumental terms.) Alternatively, the disobedients might suffer from unrealistic optimism, which might lead them to engage in civil disobedience even though the prospects for change are vanishingly small. Group influences can aggravate these effects. For its part, the authority might also be subject to outrage and from unrealistic optimism (a particular problem for those confronted with civil disobedience), leading to responses that play directly into the hands of the disobedients.

Bounded rationality on the part of voters and leaders can also encourage civil disobedience. If voters, for example, ignore problems unless they are made salient, then civil

[1] In a sense, civil disobedience serves to generate "hatred" against the leader who engages in harsh repression as in Glaeser (1995).

[2] We bracket extreme circumstances.

disobedience can serve to generate salience. If leaders are prone to excessively heavy responses which make protests more effective with voters, then this will also increase the appeal of civil disobedience to those who want a change in regime.

A great deal of empirical work would be necessary to evaluate these speculations, but we offer some brief remarks on the possibilities. We also bring the analysis to bear on a pervasive dilemma faced by criminal prosecutors: In the face of civil disobedience, is it best to initiate proceedings, or instead to exercise prosecutorial discretion so as to leave the underlying activity unpunished?

## II. Civil Disobedience: A Historic Taxonomy

There have been law-breakers as long as there have been laws, but if all law-breaking is defined as civil disobedience then the term has no value. Our focus will be on law-breaking that is motivated by larger political purposes, or designed to effect political change. Table 1 attempts to organize the types of civil disobedience.

### A. Non-Instrumental Disobedience

Some of history's most famous examples of civil disobedience were fundamentally non-instrumental. Antigone's decision to disobey the law of King Creon and bury her brother is surely civil disobedience, but it is motivated by conscience alone. Since Sophocles ascribes her decision to a desire to follow her conscience rather than the King, there is no challenge in understanding her actions.

Similarly, Thoreau's famous "Civil Disobedience" urges disobedience not because he has any faith that his actions will have an impact, but because he believes that it is immoral to support a government that allows slavery and wages war in Mexico. In a sense, Thoreau seems to believe that cooperating with the government would pollute his soul. He prefers prison time to such pollution.

While there may be interesting legal questions about such behavior – what is the appropriate punishment, if any? (Dworkin, 1967) -- modelling it would merely mean assuming that individuals have such a strong aversion to an action, for moral reasons, that they are willing to pay a large penalty to avoid taking the action. But we acknowledge that "taking a stand" can and often does often help motivate civil disobedience even when it is most unlikely to produce reform.

### B. Effective Power

We focus instead on those forms of civil disobedience that are intended to change political outcomes. (We cannot rule out the possibility that Antigone or Thoreau also hoped that their actions would have larger consequences.) The variety of such civil disobedience is enormous, ranging from disobedience in recent years in Ferguson, Missouri and Baltimore to an Indian boycott on British goods and services to the 14th Century Peasant's Revolt. One method of categorizing civil disobedience is based on *the effective power of the group*.

1. Revolution. If the group is mighty enough to produce a real challenge to the military power of the state, then any uprising offers at least the threat of out-and-out revolution. In some cases, the disobedience begins directly as a rebellion, such as the Peasant's Revolt, where Wat Tyler led a mass of ordinary Englishman to the capital, attacking prisons and legal structures. In other cases, an uprising begins peacefully, but as the crowd expands, violent confrontation ensues either accidentally or at the direction of the uprising's leaders.

The February Revolution in St. Petersburg turned from peaceful disobedience to violent conflict in an apparently haphazard, unplanned function. Its ultimate success hinged on the mutiny of the Tsar's soldiers. The French Revolution begins with the relatively peaceful actions of the Estates General, but edged towards violent conflict with the formation of the National Assembly and the defection of formerly royal soldiers to the National Guard. The storming of the Bastille was the planned toppling of royal authority in central Paris.

Both events remind us that the success of a popular uprising often hinges more on a psychology than on real or apparent military might. The central question in many uprisings is whether soldiers are willing to obey orders and fire on unarmed disobedients. The unwillingness of Egyptian soldiers to fire on the crowds in Tahrir Square marked the obvious end of the Mubarak regime.

2. *Ensuring bargaining*. A second form of instrumental civil disobedience involves causing non-violent pain to political leaders in the hopes of bringing them to the bargaining table. In these cases, the disobedients are either too weak to effectively rebel or choose, at least temporarily, to avoid the downsides of violent conflict. Colonialists used boycotts of British goods after the stamp tax of 1765, which may have helped induce the repeal of that act. Boycotts were also used to disobey the Townsend Acts of 1767. Gandhi employed a similar tactic in 1921, with non-cooperation and the Swadeshi policy, which represented a boycott of British goods. In 1931, the Viceroy Lord Irwin agreed to a series of Gandhi's demand in exchange for an end to the non-cooperation.

General strikes, aimed at governments more than companies, also fit within this middle category of civil disobedience. The British General Strike of 1926 aimed more at getting support for coal miners from Stanley Baldwin's government than it did at moving the mine owners themselves. The Austrian General Strike of 1950 may have been intended to bring Communists to power.

Civil disobedience as a bargaining tool seems, to us at least, to be so similar to a labor strike that there is little need for new theory. The economic literature on labor unions as strikes, including Ashenfelter and Johnson (1969), Farber (1978), Jun (1989) and Fernandez and Glazer (1991), is copious and filled with insight. The models from the strike and bargaining literature can be readily used to understand the economics of civil disobedience as a bargaining tactic.

*3. Provoking authority and shifting public opinion.* We focus instead on civil disobedience by groups that are too weak to effectively generate either a revolution or to cause significant direct harm to the political leadership over a sustained period. Our principal concern is civil disobedience that provokes a forceful (hostile) response from authority, but disobedience can also produce reform in the absence of that response, and what we say touches on that phenomenon as well. In some cases, disobedients may hope that the sheer size of their protest will sway public opinion.

The civil disobedience that occurred in October 2014 after the killing of Michael Brown in Ferguson, Missouri, certainly posed no direct military threat to government of either the U.S. or Missouri. Neither the state nor the city governments were hampered materially by the disobedience itself, which involved largely nighttime conflicts. Moreover, it was not obvious that the disobedient had leaders who could effectively bargain or commit to perpetuate the disobedience until demands were met. Nonetheless, the disobedient said that they were hoping to change outcomes in the short run (indict the policemen) and in the long run (change police behavior towards African-Americans). They certainly did not say that they were just expressing themselves.

A similar example followed the arrested on April 12, 2015 of Freddie Gray, who died a week later in police custody at the University of Maryland Shock Trauma Center as a result a severed spinal cord.[3] For nearly two weeks following Gray's arrest and injury, the Baltimore Police department maintained that the police involved in his arrest had not used excessive force.[4] Peaceful protests against the lack of transparency and accountability around Gray's arrest began prior to Gray's death, but escalated in size and disruption following his death and the April 24th acknowledgement by the Baltimore Police Commissioner that Gray had not been given necessary medical attention and had not been wearing a seatbelt while being transported to the police station.[5]

The signal worked: On May 1st, four days after the worst of the rioting, the short-term goals of the protestors were met. The state's attorney for Baltimore, Marilyn Mosby, announced that charges ranging from reckless endangerment to second degree murder would be brought against

[3] Eric Ortiz, "Freddie Gray: From Baltimore Arrest to Protests, a Timeline of the Case" NBC News, May 1, 2015. Retrieved at: http://www.nbcnews.com/storyline/baltimore-unrest/timeline-freddie-gray-case-arrest-protests-n351156
[4] Id.
[5] Id.

all six officers involved in Gray's arrest.[6] In outlining the charges, Mosby described the events surrounding Gray's arrest and the officers' actions that resulted in their criminal liability for Gray's death, providing the transparency and accountability sought by protestors.

Though the destruction and violence of the rioters resulted in unsympathetic media coverage, the police response to the disobedience revealed certain elements of the Baltimore law enforcement and criminal justice system that protestors sought to expose in pursuit of their long-term objectives. The widespread arrests were criticized as failing to distinguish between peaceful protestors and destructive rioters, particularly as peaceful demonstrators were arrested days after the initial violence for violating the curfew that remained in place.[7] Human rights organizations criticized the exorbitant bail set for arrested protestors as well as the governor's order allowing those arrested during the protests to be held longer than 24 hours without charges or bail.[8] By April 29, roughly half of those arrested during the violence of April 27th were released without bail or charges amid criticism that they had been illegally detained.[9]

Such forms of civil disobedience, like non-violent disobedience in many democracies, seems aimed at producing political reform by changing public opinion more broadly, which may change policy either by voting a leader out of office or by prodding that leader to change course, perhaps in fear of being ousted. In response to the disobedience, Governor Nixon did switch policing duties for Ferguson from the St. Louis County Police to Missouri State Highway Patrol, which can be seen as something of a victory, albeit a modest one, for the disobedient.

The attempt to shape public opinion is significant even in larger forms of civil disobedience. Gandhi's 1930 Salt March was far more politically powerful as a symbol of the Indian desire for self-rule than as a reduction in the revenues of British salt monopoly. As Gandhi himself wrote "Satyagraha," or nonviolent resistance "is a process of education public opinion such that it covers all the elements of society and in the end makes itself irresistible."

Most disobedience marches are meant to alter public opinion, such as the 1963 Great March on Washington. There was a debate during the Great March about how much to focus on inconveniencing political leaders (by shutting down Washington) and how much to focus on just demonstrating the enormous scope of unrest. The mere threat of a march on Washington in 1941 had prodded President Roosevelt into desegregating the war industries. The 1894 march by

---

[6] Ortiz.
[7] "Baltimore protestors arrested defying curfew" ABC News, May 1 2015, http://abcnews.go.com/US/baltimore-protesters-arrested-defying-curfew/story?id=30748375
[8] Justin Fenton, "Judge Oks Hogan's order to extend hold on riot suspects" Baltimore Sun, May 4, 2015, Retrieved at: http://www.baltimoresun.com/news/maryland/baltimore-riots/bs-md-ci-hogan-order-challenge-denied-20150504-story.html#page=1
[9] Luke Broadwater, et. Al., "Half of those arrested in riot released without charges", Baltimore Sun, May 1, 2015, Retrieved at: http://www.baltimoresun.com/news/maryland/baltimore-city/bs-md-ci-riot-released-20150429-story.html#page=1

Coxey's Army and the 1932 Bonus March were notably less successful, partially because they faced far more hostility in Washington, including serious armed opposition in 1932.

Gandhi's thinking on non-violence seems to have been partially shaped by Tolstoy's "The Kingdom of Good is Within You," which is something of a bridge between non-instrumental and instrumental non-violence. The bulk of Tolstoy's writing supports non-violence for largely non-instrumental moral reasons, but in Chapter IX, he expounds upon the political power of non-violent disobedience. He writes that "To punish men for refusing to act against their conscience the government must renounce all claim to good sense and benevolence," which eliminates the moral authority of government, because "they assure people that they only rule in the name of good sense and benevolence." According to Tolstoy, non-violence is so powerful because "authorities are in such a defenseless position before men who advocate Christianity, that but little is necessary to overthrow this sovereign power which seems so powerful."

Gandhi read and revered Tolstoy and urged civil resisters to "joyfully suffer even until death," presumably because "those who die unresistently are likely to still the hand of violence by their wholly innocent sacrifice." The innocence of the nonviolent resisters was critical, and he urged them to "put up with assaults from the opponent, never retaliate," "protect" officials from insult or attack, and "behave courteously towards prison officials." Such policies would make little sense if civil disobedience was the political equivalent of a labor strike, the goal of which is to inflict economic pain on your opponent. Gandhi may have advocated cheerful suffering for primarily moral reasons, but it was also effective politics that energized his supporters and reduced British support for the Raj.

Independent of Gandhi and Tolstoy, there was a strong traditional of nonviolence within the U.S. and elsewhere. Thoreau was an early proponent, and Quakers served as a continuing source of advocacy for nonviolence, through organizations like the American Friends Service Committee, which helped train Bayard Rustin and helped funded Martin Luther King's 1959 visit to India to study Gandhism. Both Rustin and King studied Gandhi. His success – achieving independence through non-violent disobedience—greatly increased the appeal of this approach to a wider range of social actors, particularly those in the Civil Rights Movement in the United States.

The success of the Civil Rights Movement during the 1960s had little in common with either violent revolution or a general strike. The southern states, which were presumably paying the direct cost of Freedom Rides, sit-ins and other disobedience, never independently softened their policies. Non-violent civil rights disobedience were effectively in generating non-Southern support for civil rights. The images of non-violent young disobedients facing the dogs and hoses of southern authority and the deaths of northern activists in the south helped change the political atmosphere outside the south.

The model that follows is meant to describe events of that kind, where civil disobedience causes political change by altering political views. The model is largely rational, where both

disobedience and the repression of disobedience provide new information to the median voter. We later discuss the role of psychological factors for the disobedient and for leaders, involving anger, salience, and unrealistic optimism.

## III. Instrumental Civil Disobedience

We now formally model the decision to engage in civil disobedience. The core idea that drives the model is that disobedience can create political change by generating information revelation. The model begins with the decision of private individuals to engage in civil disobedience. A political leader then decides whether or not to repress this disobedience. Voters then absorb what has happened and the decide whether to re-elect or oust the leader. We now go through those steps and describe the model's formal assumption.

Acts of civil disobedience, both in the paper and in the real world, have at least two critical dimensions: the number of disobedients and the damage done by each act of disobedience. Massing illegally on a public space on a Sunday does little damage, but may involve a large number of disobedients. Blowing up a Federal Post Office building does a great deal of damage even though it involves only a small number of disobedients.

To capture this distinction, we will separate out the disobedients into a planner and followers. The planner will choose the nature of the civil disobedience thinking strategically about how that will impact the number of people who act and the influence of the mass action. The disobedients themselves are too small to have an individual influence and therefore will be guided by fundamentally non-instrumental motives.

We make the simplifying assumption that the total damage D equals the number of participants in the act of civil disobedience, N, times the damage per disobedient d. We assume that the individual participants can decide about whether they want to participate in the action, but that there is a unitary actor – the disobedience planner—deciding on the nature of the disobedience and hence the value of d. This value of d may then impact the number of participants in the protest.

The disobedience planner's choice may have major impacts on the public sphere, and therefore we assume that choice is motivated primarily by instrumental factors. By contrast, we assume that each potential disobedient is only a tiny part of the total movement, and therefore the disobedients are motivated by non-instrumental reasons. We allow for the possibility that protesters prefer to be involved in larger events. Nonetheless, the nature of the civil disobedience may influence the number of protesters who attend, both because it influences their desire to participate and because it will determine the response of civil leadership.

These assumptions are simplifications. In many cases, individual protesters decide both on participation and how much damage to do. In other cases, there is far more central planning of protests, as there was during the Civil Rights movement and by Gandhi.

The protest will have influence on the election if it reveals information about the leader. We will assume that leaders differ along two dimensions: generic toughness and hostility to the unhappy group that may engage in civil disobedience. Moreover, we will treat this dimension as binary, so that a leader can be either tough or benign, in terms of general toughness, and can be either hostile or neutral, in his attitude towards the group that is engaging in disobedience. Only the leader knows whether he is tough or benign. The potentially disobedient group knows whether the leader is hostile or neutral towards them. One goal of civil disobedience may be to signal the unhappiness of a group. A second goal may be to get the leader to reveal his level of toughness.

The primary effect of the leader being tough is that he will have a lower cost of repressing disobedience than if the leader is benign. This cost is largely meant to be psychic, perhaps because benign leaders have more empathy towards the disobedient group. We also assume that voters' preferences for re-electing the leader may depend on whether the leader is thought to be tough or benign.

The leaders' hostility to the out-group only matters to that group. If the leader is hostile to the group, then the members of that group have a greater taste for civil disobedience and the planner of the protest cares more about seeing the leader lose his position.

Since there are two dimensions and two possibilities for each dimension, there are four possible types of leader. The probabilities of the four possibilities are summarized in the following table:

| Probabilities of Leader Type | Leader is hostile to disobedient group | Leader is neutral to disobedient group | Total |
|---|---|---|---|
| Leader is Tough | $\sigma\mu p_0$ | $(1-\sigma\mu)p_0$ | $p_0$ |
| Leader is Benign | $(1-\sigma p_0)\mu$ | $1-\mu-p_0+\sigma\mu p_0$ | $1-p_0$ |
| Total | $\mu$ | $1-\mu$ | |

The unconditional probability of the leader being tough, rather than benign, is $p_0$. The unconditional probability of the leader being hostile, rather than neutral, to the disobedient group is $\mu$. The correlation between the two events is captured by the parameter $\sigma$. When $\sigma = 1$, the

two events are unrelated and the group has essentially no information to reveal about the leader. When $\sigma > 1$, then the leader is more likely to be tough, if he is hostile to the disobedient group.

The disobedience planners' welfare if the leader is ousted equals $B_j \in \{B_H, B_N\}$, where $B_H$ is the welfare if the leader is hostile and $B_N$ is the welfare if the leader is neutral, with $B_H > B_N$. The protest is planned maximizing the expected value of $B_j$ minus $\xi d$, where $\xi$ is arbitrarily small but still strictly greater than zero. The role of $\xi$ is just to ensure that if there are many different types of protest that yield identical political outcomes, the protest planner will always choose the type of protest that has the lowest value of d.

The disobedients themselves may also hope that a tough leader is removed, but we need not model that aspect of their utility, since their individual actions will not impact the probability that the leader is removed. Their welfare is normalized to zero if they do not participate in civil disobedience. If they join in the action, their welfare from disobedience will be $b_j + \varepsilon_k$ if the disobedience is not repressed or $b_j + \varepsilon_k - c(N)$ if the disobedience is repressed. Again, $b_j \in \{b_H, b_N\}$, where $b_H$ is the benefit from disobedience if the leader is hostile and $b_N$ is the benefit if the leader is neutral, with $b_H > b_N$. These benefits and costs are largely psychic.

We assume $C(N) \geq 0$, $C'(N) \leq 0$, $C"(N) \geq 0$ and $lim_{N\to\infty} C(N) = 0$. The core idea is that the cost of being repressed are positive, but get smaller as the size of the protest increases, reflecting the fact that the probability of arrest is small in a larger crowd as in DiPasquale and Glaeser (1997). However, the impact of crowd size has diminishing returns, and ultimately the probability of arrest goes to zero and N gets arbitrarily large. We assume that the maximum value of $\varepsilon_k$ is denoted $\varepsilon_{Max}$ and we typically assume that this value is finite. We assume that there is a symmetric, single-peaked distribution of $\varepsilon$, with a mean and median of zero that is described by a cumulative distribution $G(\varepsilon)$ and a single-peaked density $g(\varepsilon)$, hence $g'(\varepsilon) \geq 0$ and $g"(\varepsilon) \leq 0$ for $\varepsilon < 0$ and $g'(\varepsilon) \leq 0$ and $g"(\varepsilon) \geq 0$ for $\varepsilon > 0$.

The total population of potential disobedients is denoted Q. Therefore size of the protest will equal $(1 - G(-b_j))Q$ if it is known that the protest will not be suppressed. If it is known that the protest will be suppressed, then the size of the protest is a fixed point of the equation $\left(1 - G\left(c(N) - b_j\right)\right) Q = N$, which will always be less than the size of the protest if it is known that the protest will be not be repressed.

Our bifurcation between planner and individual disobedient means that there we have separated out the two functions of most disobedience: regime change (which is the interest of the planner) and non-instrumental expression of unhappiness (which is the interest of the individual disobedient).

After the decision about the nature of the protest and the amount of protest, the leader will have the choice of two actions: repression or accommodation. The critical assumption is just that

there is one response which harms the disobedients more and second that harms them less. In principle, the tougher response could include prosecution or police brutality. The lesser response could include changing policies in line with the disobedients' requests or merely benign neglect.

The leader's welfare equals V, a continuation value, times the probability that he remains in office, minus D if he accomodates or $K_l$ if he represses, where $K_l = \varphi K$ if the leader is tough and K if the leader is benign. The level of D represents the inconveniences and embarrassment to the leader if disobedients have seized public or private spaces, or whatever other downsides are linked to the disobedience. The costs $K_i$ reflect the embarrassment and inconvenience of using the forces of the state to clear the streets and restore order. These costs of repression occur at the time when the disorder is suppressed; they are distinct from the electoral consequences of dealing with the disorder.

The value of V reflects the benefits of leadership, the extent to which the leader values the future, and the probability that the leader will be ousted for some other reason in the future.

The leader is then be accountable to voters, or possibly, someone further up in a hierarchy. Disobedience in democracies typically seek to influence the electorate, but in more dictatorial regimes, disobedience of a regional leader may be seeking to have him replaced by central leadership. Even dictators often rely upon tacit support from elites and the army, and the opinions of these groups could be shaped by widepread disobedience. Voters may have their opinions altered either by the disobedience or by the response of the leader.

We have assumed that leaders have heterogeneous costs of repressing disorder but the cost of accommodating are homogeneous, but this is largely irrelevant. The important assumption for the model is that there is heterogeneity in the difference in leader welfare between accommodating and repressing.

In the third period, voters choose whether to re-elect the leader. We assume that all voters vote and vote to reelect the leader if and only if the net benefit from re-electing the leader is positive. For voter i, the net benefit for re-electing the leader is $\theta_L - \gamma p_V + \xi + \varepsilon_i$. They will vote to re-elect the leader whenever this quantity is positive. The first term, $\theta_L$, is a constant that reflects the leader's core appeal, including his charisma, past track record and other attributes as a leader. The second term reflects the beliefs about the leaders' character, where $p_V$, is the voters' belief after the disobedience has occurred that the leader is tough. If $\gamma$ is positive, then voters will be less likely to support a leader who is perceived as tough. If $\gamma$ is negative, then the opposite is true and voters prefer law and order.

The last two terms, $\xi$ and $\varepsilon_i$, are noise terms that equal zero in expectation. The term $\xi$ is a common shock that impacts all voters but is not known at the time of the disobedience. The cumulative distribution of $\xi$ is described by a function F(.). The $\varepsilon_i$ term reflects individual tastes that also have mean zero. The leader is re-elected if a majority of voters support the

leader's re-election if and only if $\theta_L - \gamma p_V + \xi > 0$. The probability of the leader being ousted is therefore $F(\gamma p_V - \theta_L)$. We let $p_D$ denote the probability that the leader is tough conditional upon disobedience taking place. now adopt the notation: $\vartheta_1 = 1 - F(\gamma - \theta_L)$, $\vartheta_0 = 1 - F(-\theta_L)$ and $\vartheta_D = 1 - F(\gamma p_D - \theta_L)$, where $\vartheta_1 < \vartheta_D < \vartheta_0$ as long as $\gamma > 0$. These are the probabilities of winning re-election condition upon the beliefs of the voters.

From our perspective, voters are essentially just a machine for turning beliefs about the leader into an outcome that both the leader and the disobedience planner care about. Our final key assumption concerns belief formation by voters off the equilibrium path.

As discussed already, tough leaders have a lower cost f engaging in repression. We will assume that voters will always interpret repression as being an indication that leader is more likely to be tough. More formally, following the logic of the D1 Refinement discussed by Banks and Sobel (1987) and Cho and Kreps (1987), we assume that if voters expect all leaders to repress, then they will believe that a leader who does not repress is benign. If voters expect all leaders to do nothing, then they will believe that a leader who responds harshly is tough.

*The Leader's Decision*

As we have already specified the behavior of voters in period 3, we now proceed recursively to period 2 and turn to the decision of a leader who is facing disorder of size D. The leader's behavior is response to that disorder will then shape the decisions of the disobedients in the first period.

The leader's welfare equals $\big(1 - F(\gamma p_V - \theta_L)\big)V$ minus the costs of either repressing disorder or doing nothing. The leader's decision is complicated because his action may signal his type. Repressing or accommodating acts of civil disobedience become a complicated signal indicating whether the leader is benign or tough.

The equilibria of this model can take three forms: pooling, separating and semi-pooling. In a pooling equilibrium, both types of leaders take the same action. In a separating equilibrium, they take different actions. In a semi-pooling equilibrium, leaders of one type randomize between actions while leaders of the other type usually take a single action.

We also let $p_D$ denote the probability assigned by voters to the leader being malign conditional upon disorder occurring, which may be different from $p_0$.

We focus on the case where $\gamma$ is positive, so that leaders wish to appear as benign, but we also discuss the largely symmetric case when $\gamma$ is negative.

*Proposition 1:* (i) If $K > \varphi K > D$ and $\gamma > 0$, then neither type of leader will suppress disorder harshly.

(ii) If $K > D > \varphi K$ and $\gamma > 0$,, then there exists a value of V, denoted $\underline{V}$, where if $V < \underline{V}$, h all tough leaders repress and benign leaders do nothing, and there also exists a second greater value of V, denoted $\overline{V}$, where if $V > \overline{V}$, all leaders do nothing. If $\underline{V}, < V < \overline{V}$, then tough leaders randomize between the two responses, while benign leaders do nothing.

(iii) If $K > D > \varphi K$ and $\gamma < 0$,, then there exists a value of V, denoted $\underline{V}$, where if $V < \underline{V}$, h all tough leaders repress and benign leaders do nothing, and there also exists a second greater value of V, denoted $\overline{V}$, where if $V > \overline{V}$, all leaders repress. If $\underline{V}, < V < \overline{V}$, then benign leaders randomize between the two responses, while tough leaders always repress.

(iv) If $D > K > \varphi K$ and $\gamma > 0$, then when V is low, all leaders repress and if V is sufficiently high, all leaders do nothing. For intermediate levels of V, there can be three equilibria: one in which all leaders repress, one in which benign types randomize between the actions, while tough leaders repress, and one that depending on the value of V, can include involves separating, semi-pooling or pooling.

(v) If $D > K > \varphi K$ and $\gamma < 0$, then all leaders will repress.

Proposition 1 details the basic predictions of the model about behavior of the leader and Table 3 reports the parameter values under which different equilibria can occur. If $K > \varphi K > D$, then the costs of repression are greater than the costs of doing nothing for both types of leader, and both types of leaders do nothing. If voters prefer benign leaders, then there is never any strategic reason to be harsh, given our assumption that voters prefer benign leaders. As such, if the disorder is sufficiently mild then there is no reason to pay the costs of beating it down.

This result would change if voters actually preferred tough leaders. In that case, for high enough values of V, tough leaders would start to repress, even if their own preferences favored leniency, in order to signal their toughness. We will not explicitly deal with this case, primarily because civil disobedience is particularly inexplicable if voters have a taste for leaders who are tough enough to harshly repress civil disobedience.

The middle case, in which $K > D > \varphi K$, is the most interesting. For those parameter values, benign leaders would intrinsically prefer doing nothing while tough leaders would like to repress. Politics, however, pushes tough leaders towards tolerance if $\gamma > 0$. If V is sufficiently low, so that career concerns are relatively unimportant, then tough leaders act tough and ignore the political consequences. For higher values of V, however, tough leaders start imitating the benign leaders and do nothing. This can be understood as capturing the dictator who allows the disobedience to continue, ignoring his core instincts, because he wants to show a good face to the world. At the highest levels of V, harsh punishment disappears altogether, because tough leaders are really desperate to keep their jobs.

These results would reverse if $\gamma < 0$, and voters wanted tough leaders who would repress disorder. In that case, again assuming that $K > D > \varphi K$, leaders continue to act their type when V is low. But if $\gamma < 0$, as V rises, then benign leaders increasingly start acting tough and for very high levels of V, both types of leaders repress any disobedience. Oddly, in this case, disobedience stands the best chance of leading to the ousting of the most benign leaders, which should presumably deter protesting. If $\gamma < 0$, and $D > K > \varphi K$, then disobedience always engenders repression. We have avoided characterizing the case where $\gamma < 0$ and $K > \varphi K > D$, because it is complex and in our opinion, unlikely to be all that relevant.

Our focus on the case where voters prefer benign leaders to tough leaders, does not imply that we think that this is the norm in U.S. history. In many cases, voters seem to prefer law-and-order candidates to accomodationists. Naturally, this will create the possibility that disobedience will increase the probability of ousting a benign leaders—a boomerang effect-- where disobedience entrenches the tough and ousts the mild. A classic example of that boomerang effect was when the 1967 Detroit riot discredited liberal Mayor Jerome Cavanagh and led to the election of the far more conservative Roman Gribbs.

Returning to the case where $\gamma > 0$, we next consider the setting where $D > K > \varphi K$, which is fairly complex. When V is very low, then both leaders act tough, which is their preferred action in the absence of career concerns. When V is very high, the both types of leaders act leniently, despite the fact that neither is innately interested in being lenient. This outcome is related to the problem of wasteful signaling first highlighted by Spence (1973). Both types of leaders are doing something that neither wants to do, because if they don't, then voters will think that they are tough. In this case, the interests of the wider public may be different from the interests of the leaders and as such, the signaling done by tolerating disorder could be beneficial.

For intermediate levels of V, there will be exactly three equilibriums. If V is somewhat greater than $\frac{D-K}{\vartheta_0-\vartheta_1}$, then three equilibrium will include one equilibrium in which both types suppress, one equilibrium in which only malign types suppress and one equilibrium in which malign types randomize between repression and doing nothing. We define $\vartheta_0$ as the posterior belief that the leader is tough condition upon the leader. This multiplicity exists because if all types are repressing, then benign types gain less by switching to doing nothing. If only tough types are repressing, then benign types don't want to switch to suppression because then voters will think that they are surely malign. Likewise, there is an equilibrium between these other two in which all some but not all benign types act leniently. In this case, the gain in reputation from acting leniently is less than in the separating equilibrium but more than in the pooling equilibrium, and exactly enough to make the benign types indifferent between the two actions.

As V rises, the configuration of equilibria changes slightly. For higher values of V, the separating equilibria disappears and becomes another semi-pooling equilibrium. Eventually, as

V becomes sufficiently high, the number of equilibria drops from three to one. The basic intuition when $D > K > \varphi K$ is not so different than when $K > D > \varphi K$, but the predictions are more complicated, and we will focus on comparative statics when $K > D > \varphi K$. The next proposition provides comparative statics assuming that condition holds:

*Proposition 2:* When $K > D > \varphi K$ and $\gamma > 0$, then $\underline{V}$ (the highest value of V for which all tough leaders repress) increases with D, declines with K, $\varphi$, and $\gamma$ and increases with $\theta_L$ if and only if $f(-\theta_L) < f(\gamma - \theta_L)$. The value of $\overline{V}$ (the lowest value of V for which all tough leaders do nothing) increases with $D$ and $p_D$ and decreases with $\varphi$ and K. The value of $\overline{V}$ increases with $\theta_L$ if and only if $f(\gamma p_D - \theta_L) < f(\gamma - \theta_L)$ and decreases with $\gamma$ if and only if $p_D f(\gamma p_D - \theta_L) < f(\gamma - \theta_L)$. If $\overline{V} > V > \underline{V}$ , then the share of tough leaders who do nothing is falling with D and rising with $\varphi$, K and V.

The logic of the proposition is relatively straightforward. When V is low, then the benign types are tolerant and the tough leaders repress. The range of values of V for which this occurs expands with D because larger values of D, the costs of disorder to the leader, make it less attractive for tough leaders to imitate the benign leaders and do nothing. Lower values of K and $\varphi$ also expand the range of values of V where tough leaders repress because they reduce the costs of repression. A higher value of $\gamma$ will increase the desire to be seen as benign and therefore make it more attractive for tough leaders to want to appear to be benign.

If the leader is innately stronger, with a higher value of $\theta_L$, then this will expand the range of values for which repression is attractive to tough types as long $f(-\theta_L)$ is less than $f(\gamma - \theta_L)$. If $f(\gamma - \theta_L)$ is higher, this means that charisma is more valuable with voters if the leader is thought to be malign, and therefore complements suppression.

When V is greater than $\overline{V}$, then the tough leaders do nothing in order to appear benign. This cutoff should be interpreted as one measure of how attractive it is for tough leaders to imitate the benign. This lower bound is rising with D and falling with $\varphi$ and K because $D - \varphi K$ determine the net cost, for the tough types, of doing nothing. In this pooling equilibrium, a higher value of $p_D$ will lead voters to believe that tolerant behavior still comes from a malign leader. This will make pooling less attractive, holding V constant. The lower bound rises with $\theta_L$ as long as $f(\gamma p_D - \theta_L)$ is less than $f(\gamma - \theta_L)$, which should be interpreted as meaning that charisma complements being tough.

Perhaps most importantly, this lower bound, $\overline{V}$, decreases with $\gamma$ if and only if $p_D f(\gamma p_D - \theta_L)$ is less than $f(\gamma - \theta_L)$. This condition will hold as long as the densities are similar, which seems reasonable. Higher values of $\gamma$ mean that the population really wants a benign leader, and unsurprisingly, this will make it more likely that the tough will imitate more tolerant leaders. Conversely, as $\gamma$ goes to zero, this cutoff goes to infinity. If the population doesn't value

tolerance much, then there is essentially no case in which the tough will imitate the benign. If $\gamma$ were negative, then with high values of V, the benign would imitate the malign.

The last set of comparative statics concern the share of tough types who choose to imitate the benign in the semi-pooling equilibrium. This parameter is important because it determines the expected amount of suppression. The share of malign leaders who are tolerant falls with D. Unsurprisingly higher costs of disorder lead to more suppression. Higher values of $\varphi$ or K, which capture the costs of suppression, lead to less suppression. Higher values of V, which implies a greater weight put on re-election, also leads to more imitation of the benign.

*The Choice of Civil Disobedience by Individual Protesters*

We now turn to the choice of disobedience. There are actually two separate choices to consider. The decision of the individual protester and the decision of the protest planner. Since the planner moves first, we must continue to solve the model recursively, first describing the behavior of the protester for a given value of d.

From the protesters' perspective there are three possibilities. First, if the protest will not be repressed then the protesters will participate as long as $b_j + \varepsilon_k > 0$. We will always assume that $b_H + \varepsilon_{Max} > 0$, so that there is some potential for protest, but that $b_H + \varepsilon_{Median} < 0$, so that the protesters always represent a minority of the total group. As a general point, the size of the protest will generically fully reveal the value of $b_j$. As a result, the protest planner will be able to transmit the information about the state of his group's unhappiness just by having a protest. It doesn't need to be large.

If it is known that the protest will be suppressed with probability $\pi$, then the size of the protest is a fixed point of the equation $1 - G\big(\pi c(N) - b_j\big) - \frac{N}{Q} = 0$. There can easily be multiple equilibria of this function. For example, if $\pi = 1$ and $G\big(c(0) - b_j\big) = 1$, and there exists a point which $\frac{Q-N}{Q} > G\big(c(N) - b_j\big)$, then there can readily be three equilibria. The first has no disobedience. In the second, there is some disobedience but the equilibrium is unstable in the casual sense that a slight increase in the number of disobedients would cause the returns from being disobedient to rise. In the third, there is a stable and significant amount of disobedience.

We make a number of simplifying assumptions. First, we assume that c(.) is convex, reflecting the decreasing returns to having fellow protesters. Second, we assume that G(.) is single peaked at the median. This will reduce the possible types of multiplicity of equilibrium.

*Proposition 3:* If $\frac{\varepsilon_{max}+b_j}{c(0)} > \pi$, then there exists a unique equilibrium with a positive level of disobedience, and the level of disobedience is rising with Q and $b_j$ and falling with $\pi$. If $\frac{1}{Q} > -g(\varepsilon_{max})\big(\varepsilon_{max} + b_j\big)\frac{c\prime(0)}{c(0)}$, then there is no equilibrium with positive disobedience for

$\frac{\varepsilon_{max}+b_j}{c(0)} < \pi$. If $\frac{1}{Q} < -g(\varepsilon_{max})(\varepsilon_{max}+b_j)\frac{c\prime(0)}{c(0)}$, then for all values of Q there will exist a unique value of $\pi$, which determines the maximum level of repression, denoted $\pi_{max}$ at which disobedience can occur. The value of $\pi_{max}$ is rising with Q and $b_j$. If $\pi$ lies between $\frac{\varepsilon_{max}+b_j}{c(0)}$ and $\pi_{max}$, then there are multiple equilibria: one with no disobedience, and two with disobedience. The share of potential protesters who are disobedient in the equilibrium with more disobedience is rising with Q and falling with $\pi$.

Proposition 3 characterizes the behavior of potential disobedients, which is determined by the probability of repression, the intensity of preferences, and the size of the group. When the probability of repression is low (less than $\frac{\varepsilon_{max}+b_j}{c(0)}$) then there is a unique equilibrium and it is well behaved. When the size of the group is bigger, a greater share of the group will be disobedient, except in the limit where the probability of repression is zero. Group size matters because larger groups mean lower costs of repression to each person who is disobedient.

The probability of repression increases the costs of disobedience and causes the size of the protest to decline. Intensity of preferences also matters, unsurprisingly, since groups with strong tastes will have more protesters holding the probability of repression constant.

One important corollary of this proposition is that the size of the protest completely reveals the depth of protestor preferences. In any equilibrium with a positive amount of protesting, the strength of preferences will be completely revealing. Signaling the group's preferences can therefore be done quite cheaply, with an arbitrarily non-threatening protest. Since protesters don't behave strategically on the individual level, their behavior is always revealing. This is a big advantage for protests that are formed by voluntary attendance of the many, as opposed to concerted acts of the few.

If $\frac{\varepsilon_{max}+b_j}{c(0)} < \pi$, and if Q is small, then the only equilibrium involves no disobedience at all. If Q is larger than there will exist multiple equilibrium. In one equilibrium, there is no disobedience. There is a second equilibrium that is unstable in the casual sense that if slightly more people were disobedient the net returns to disobedience would rise and even more people would be disobedient. There is a third equilibrium with a higher level of disobedience and at that level there is stability. The same comparative statics apply for the level of disobedience in that third equilibrium.

The proposition highlights that for a given level of Q and the depth of preferences there is a maximum level of $\pi$ that permits disobedience to occur. When Q is high, then an equilibrium disobedience can exist even if the probability of repression is one. If Q is low, then such an equilibrium does not exist. Protesters will not be willing to show up in they are sufficiently sure that there will be repression.

The proposition illustrates the constraints faced by the planners of civil disobedience. They need to ensure that protesters participate. Participation can be sure either if (1) the probability of repression is small, (2) the taste for disobedience is high or (3) the size of the community is large enough so that the costs of repression to the individual protester are small. Essentially, either strong tastes or a big community can free a protester to take actions that are really harmful to the existing regime.

We will deal with the multiple equilibria issue by assuming that the planner always has the power to select the equilibrium that is chosen by coordinating the activities of his people. This will mean that if a given set of parameter values admits an equilibrium with disobedience that will be the equilibrium that will occur. This is a significant assumption, but it does capture the organizing role of the planner and it seems a reasonable a way of dealing with multiple equilibria. We do not, however, allow the planner to choose which equilibrium occurs when the multiplicity is on the part of the political leader.

*The Choice of the Disobedience Planner*

Just as there are two types of leaders, there are effectively two types of disobedient planners: those who know that the leader has been hostile to them and those who know that the leader has not been hostile to their group. That information is private and the planner wants to signal that information, but it will be signaled as long as he can organize any sort of disobedience at all. Since the individual disobedients are non-strategic, the size of the action will automatically reveal the state of their minds. This is not a feature of two states, the value of $b_j$ is revealed as long as there is any disobedience.

Lemma 1: The disobedience planner always prefer the equilibrium with an infinitesimal level of "d" and no repression (the epsilon protest) to any other equilibrium in which all types of leaders do the same thing.

Lemma 1 reflects the fact that either type of disobedience planner can costlessly reveal their type with an arbitrarily small, harmless protest that does not engender any repression We call this type of protest an epsilon-protest, which is the most non-threatening protest possible meant entirely to illustrate the number of one's supporters. This is perhaps the nature of legal, non-violent marches that disperse quickly causing little harm.

A more complex information structure would make it possible for planners to prefer more damaging events even if they were not repressed. In our model, one parameter captures the extent of out-group sentiments. If there were two parameters, one of which captured the depth of dissatisfaction and one of which captured the width of dissatisfaction then more dangerous protest might be a way of showing that anger was both wide and deep.

We now focus on the planners' decision about whether to engage in more damaging protests within the structure of the model. The only reason to take that route in our model is to elicit repression and reveal something about the political leader, since the protesters' feelings can be shown with an epsilon-protest. We assume that planners representing groups that are not hostile to the leader will not protest, and focus on the intensive margin of protest for a planner whose group is hostile to the leader. In principle, there could be a perverse situation in which planners' who represent groups that are happy with the leadership might choose to protest also, perhaps because they wanted to show voters that they were happy to help the leader keep his job. But since there is full revelation of the planners' type automatically with any protest, the situation is largely symmetric for the two groups and would be essentially the same for any distribution in the tastes of protesters and protest planners. This implies that the expected probability that the leader is tough is $\sigma p_0$ conditional upon any protest by the more harmed disobedients.

To determine the planners' welfare, we must now address the case in which there are multiple equilibria for the leaders' response to a disorder of size D. This multiplicity occurs specifically when D is greater than $\big(F(\gamma\sigma p_0 - \theta_L) - F(-\theta_L)\big)V + K$ and less than $V\big(F(\gamma - \theta_L) - F(-\theta_L)\big) + K$. For this range of values, it can be that all types of leaders repress; both types of leaders would prefer repression in the absence of career concerns. Alternatively, there can be an equilibrium in which there is semi-pooling, all tough leaders repress and some mild leaders randomize between repression and accommodation. This third equilibrium is unstable in the casual sense that if a slight higher share of mild leaders are thought to repress, then all mild leaders would choose to repress, while if a slightly higher share of mild leaders are thought tolerate, then all mild leaders tolerate. Still, its instability leads us not to consider this equilibrium.

The last possibility is that there is a sequence of equilibria, as in the case where D<K, where the leaders actions depend on parameter values. There is tolerance when D is low, then semi-pooling where some tough leaders repress and all mild leaders accommodate, separation for middling levels of D is high, semi-pooling where all tough leaders repress and then eventually total repression by all types of leaders.

We will assume that the equilibrium is known at the time that the planer is fixing his level of d, and that the resulting equilibrium generates a value of D, denoted $\widehat{D}$, which lies between $\big(F(\gamma\sigma p_0 - \theta_L) - F(-\theta_L)\big)V + K$ and $V\big(F(\gamma - \theta_L) - F(-\theta_L)\big) + K$. For values of D that are less than or equal to $\widehat{D}$, the sequence of equilibria described above occur. For value of $D > \widehat{D}$, total repression will occur. The value of $\widehat{D}$ is known by the protest planner, and it follows immediately from Lemma 1, that the protest planner will never choose a value of D that is greater than $\widehat{D}$. Total repression will be no more information to voters than total accomodation and will cost the planner more.

We first assume that Q is sufficiently high and $\xi$ is sufficiently close to zero so that $\pi_{max} = 1$, so the planner faces no restraints and can effectively chooses "D". In this case, it follows that

*Proposition 4:* If F(.) is concave over the region, $[-\theta_L, \gamma - \theta_L]$, then the protest planner will always prefer the epsilon protest. If F(.) is convex over the region, $[-\theta_L, \gamma - \theta_L]$, then: (1) if $\widehat{D} > \varphi K + V(F(\gamma - \theta_L) - F(-\theta_L))$, then the planner will set "d" so that $D = \varphi K + V(F(\gamma - \theta_L) - F(-\theta_L))$, (2) if $\varphi K + V(F(\gamma - \theta_L) - F(\gamma\sigma p_0 - \theta_L)) < \widehat{D} < \varphi K + V(F(\gamma - \theta_L) - F(-\theta_L))$, then the planner will set d so that $D = \widehat{D}$, and (3) if $\varphi K + V\big(F(\gamma - \theta_L) - F(\gamma\sigma p_0 - \theta_L)\big) > \widehat{D}$, then the planner will just choose the epsilon protest.

Proposition 4 highlights the determinants of rational disobedience when the protest planner is largely unconstrained in his ability to foment a large and painful protest and when the costs of planning a difficult protest are low. The proposition highlights the central role of the concavity of the vote distribution. Ultimately, the advantage of a high cost protest is information revelation—over and above the information that is revealed by the size of the protest itself.

We have assumed that the protest planner can, at low cost, reveal whether his group is being abused by planning an epsilon-protest and demonstrating his group's unhappiness through the size of the crowd. The potential added benefit of the protest will not be predictable to the planner. The leader may reveal his type by repressing the protest with too much determination. Since the protest planner doesn't know the leaders' type, this benefit is a roll of the dice, and like most gambles, its attractiveness depends on the concavity of the gamblers welfare function, which in this case is determined by the concavity of the vote distribution.

If the distribution of F(.) is concave, then the planner prefers an epsilon protest. If the distribution of F(.) is convex, then the planner wants to induce leaders to separate themselves. Convexity means that the planner gains more by showing the leader to be tough than he loses by showing the leader to be mild.

The distribution F(.) concerns the common shock experienced by the leader after the protest, and to build intuition, it makes sense to assume that this distribution is single peaked at zero, which implies that the distribution is convex when the probability of the leader losing is less than one-half and concave if the probability of re-election is less than one-half. The intuition then is that the protest planner will be more likely to take action against popular leaders (that have harmed his own group), against whom the marginal impact of protest is high than among unpopular leaders, who are likely to lose their positions anyway.

The planner will never choose a costly protest that has no potential information value, and that means that $\widehat{D}$ is an upper bound on the damage of the protest. This provides the point where even risk-taking planners must find a sweet spot for civil disobedience, which is high enough to elicit some repression but not high enough to elicit repression by everyone.

We now consider the more realistic case where the planner faces constraints created by his participants. We need not consider the concave case, where the planner wants an epsilon-protest. In that case, the planner can propose such a plan and the probability of repression will be zero. The risks of repression therefore will not limit the concave planner who wants an epsilon protest. The risks of repression will bother the planner who wants to cause enough trouble to induce tough leaders to reveal their types.

*Proposition 5:* If $\pi_{max} > \sigma p_0$, planners' behavior is described in Proposition 4, but if F(.) is convex over the region, $[-\theta_L, \gamma - \theta_L]$, and $\pi_{max} < \sigma p_0$ then if $\varphi K + V\big(F(\gamma - \theta_L) - F(\gamma\sigma p_0 - \theta_L)\big) > \widehat{D}$, the epsilon protest dominates, if $\varphi K - VF\left(\frac{\gamma\pi_{max}}{1-\sigma p_0+\pi_{max}} - \theta_L\right) + VF(\gamma - \theta_L) > \widehat{D} > K + V\big(F(\gamma - \theta_L) - F(\gamma\sigma p_0 - \theta_L)\big)$, then the planner will set d so that $D = \widehat{D}$, and if $VF\left(\frac{\gamma(\sigma p_0 - \pi_{max})}{1-\pi_{max}} - \theta_L\right) + VF(\gamma - \theta_L) < \widehat{D}$, then the planner will set d so that a fraction $\frac{\pi_{max}}{\sigma p_0}$ of tough leaders repress in equilibrium.

Proposition 5 emphasizes that when constraints bind, the planner chooses the level of D that generates that most separation. If $\pi_{max} > \sigma p_0$, then the planners' ideal does not run counter to the tastes of his constituents. Given the fully separation equilibrium that the planner prefers, the disobedients are still willing to turn out in force. In that case, Proposition 4 describes the planner's behavior. The more interesting case is when $\pi_{max} < \sigma p_0$, for in that case, the constituents will not protest if all tough leaders are expected to repress. The threat of repression deters them from action. The planner must therefore choose a milder protest than would be his first choice.

When $\widehat{D}$ is low, then this parameter remains is the relevant constraint. The planner then either gives up and plans an epsilon protest, or sets d so that $D = \widehat{D}$. When $\widehat{D}$ is higher, the relevant constraint becomes $\pi_{max}$ and the planner sets d so that the threat of repression does not exceed that amount. The combination of two constraints, and the interconnected choices of protest planner, activists and political leader produce a rich set of comparative statics discussed in the next proposition.

Proposition 5 also reminds us that there can a substantial difference of interest between the individual protesters and the protest planner. The planner wants to generate regime change and generating repression can help that happen. Individual protesters may well prefer not to be subjected to a policeman's nightstick. Given the preferences that we have assumed, the protester would always prefer the epsilon protest to the more aggressive actions.

Naturally, we have assumed that protesters themselves do not care about regime change. If they do, and that is quite possibly then case, then the planner may well be acting in the interests of the entire group, even if each individual protester would prefer to face a lower probability of repression.

*Proposition 6:* If the planners is unconstrained and F(.) is locally convex, D is rising with $\varphi$, K, V, $\gamma$ and falling with $\theta_L$. If the planner sets $D = \widehat{D}$, then an increase in $\widehat{D}$ will cause the size of the protest to shrink but the intensity of the protest to rise. Holding D constant, increases in $\varphi$, $K$, and V cause the intensity of disobedience to fall and number of disobedients to rise, while increases in $\theta_L$, $\sigma$ and $p_0$, cause the intensity of disobedience to rise and the number of disobedients to fall. If d is set so that a fraction $\frac{\pi_{max}}{\sigma p_0}$ of tough leaders repress, then the total amount of disorder and the level of d is rising with $\varphi$, K, V, and falling with $\theta_L, \sigma$, and $p_0$, but the number of disobedients is independent of these variables. The value of D is rising with $b_j$ and Q.

Proposition 6 describes the empirical predictions of the model about the intensity and size of rational civil disobedience. If the planner is unconstrained, then the model predicts that the level of overall disorder, D, is just high enough to induce all tough leaders to repress and no mild leaders to repress (the sweet spot), but that level can be achieved with a higher intensity of protest (d) and a lower number of protesters or a lower intensity of protest and a higher number of protesters. The planner is essentially indifferent between these options. The overall level of damage (D) is rising with $\varphi$ and K because it needs to be high enough to induce the tough leaders to pay the cost of repression. Higher values of V and $\gamma$ also induce more damaging protests, because stronger career concerns push the tough leaders to want to copy the mild leader more, and as a result the protest must be more severe to induce repression.

A higher value of $\theta_L$ reduces the size of disorder because the gains from appearing benign are less if the leader is inherently more charismatic. This means that the tough leaders have less reason to imitate mild leaders.

Some but not all of these comparative statics remain when the constraints bind. For example, if the damage is just at the upper limit set by the leaders' equilibrium ($\widehat{D}$), then other parameters do not change the overall level of disorder, which must be just low enough to avoid total repression.

*Discussion*

The model differentiates between two different types of instrumental civil disobedience. The first type as peaceful as possible and meant primarily to show the relative size of an unhappy minority. The model predicts that this type of disobedience is preferred when the protest's planners believe that they have private information that will matter to voters, about the degree of unhappiness of the out-group. The planner must also prefer the relatively predictability of a large scale, low intensity, protest to a smaller scale, higher intensity event. The preference for predictability, itself, reflects concave returns for the protester, which in the model are more likely to exist when the regime is more likely to be voted out even without the protest.

This suggests that the large scale peaceful protest is an act of confidence. The planner believes that the scale of the protest will matter and that it isn't necessary to risk violent repression. As we will discuss later, this suggests that this type of disobedience may be more common when groups believe that they are overly representative of the public as a whole.

The second type of disobedience will be smaller scale, since the risk of repression and that will be deter some potential protesters. This disobedience can occur even when the minority group recognizes that its unhappiness will have no impact on the median voter. The point of the protest is not to signal the number of unhappy group members, but rather to induce repression by the political leaders. The protest planner does not need to assume that knowledge of his group's unhappiness will generate a change in voter sentiment, but rather that there is some chance that the political regime will engage in behavior that will make the look terrible.

Even when the point of the protest is to generate a response, leaders of large groups will typically prefer milder protests when possible. As the planners' core constituency becomes smaller, his protests will become more severe, because severity is a substitute for size. A small number of terrorists can generate a massive crackdown if they do enough damage.

That example corresponds to the case in the model where the protest leader has access to a hard core of constituent who are willing to suffer any repression to act again the government. In that case, the protest planner who wants to generate his preferred outcome—the most information and the highest probability of regime change—can always get his wish. He just needs to set the damage high enough so that the tough leaders choose repression over tolerance.

The terrorist case also reminds us that over-doing the disobedience will tend to be ineffective. If the equilibrium is one in which all leaders would repress, then there is no information generated by the repression and no reason to engage in disobedience to begin with.

## IV. An Extension: Non-Persuasive Instrumental Disobedience

We now briefly discuss a slightly different setting where the point of persuasion is bargaining not regime chance. We therefore ignore the possibility of electoral consequences here and assume that has a third alternative: policy change. Policy change will cost the leader "A", and if he undertakes the policy change the disobedience disappears. Disobedience will not influence the leaders' probability of surviving in power. This might occur because there is no information to reveal or because the leader is entrenched or because the leader is term limited.

We consider this case largely archaic in U.S. politics, more relevant for the American Revolution than the Ferguson protests, although even in the American Revolution, a war was waged for elite English opinion. Changing public opinion has clearly become even more important in modern

protesting. Moreover, individual leaders in the U.S. today rarely have the ability to accommodate the demands of much disobedience.

A significant exception is that this bargaining seems quite relevant for campus protests. Typically, university leadership does have the ability to accommodate the protesters' demands. Moreover, campus leadership appears to face extraordinarily high costs of repression, presumably increases the appeal of campus civil disobedience.

If a protest occurs, the leader has the choice of changing policy, which will cost him A, ignoring the protest, which will cost D, or repressing which will cost either $K$ or $\varphi K$. If $A > Min(D, K)$, then changing policy is never the least costly option for either type of leader. Hence, the leader will never change policy and there is no benefit from civil disobedience. If $Min(D, \varphi K) > A$, then accommodation is the least costly option, even for the tough leaders. In this case, civil disobedience would be universally appealing. If $Min(D, K) > A > \varphi K$, then benign leaders will change policy in response to the disobedience will tough leaders will not. In that case, the disobedience will occur if and only if $(1 - p_0)Y > C(d)$, where Y denotes the benefit to the protest planner of accommodation.

This last case seems like a common one, where the outcome of the disobedience is uncertain and depends on the character of the leader. It may be a reasonable model for the actions of American colonists after 1763, where they engaged in civil disobedience in the hope that British leadership would accommodate their requests, which it did under the Marquess of Rockingham in 1765, but not under Lord North in the 1770s. It seems unlikely that the colonists anticipated that their actions would topple the Hanoverian dynasty. Rockingham's first term as Prime Minister was ended partially because of the repeal of the Stamp Act, but this can hardly have been a desirable outcome to the colonists.

In the non-informational archaic setting, the desired outcome for the protester is that a particular leader will change his or her policy. If the costs to the disobedience planner C(d) are independent of d,, then if $\varphi K > A$, the planner must just ensure that $dN > A$ which will ensure that both types of leaders will change policy. In this case, the disobedience will occur whenever Y>C.

If $K > A > \varphi K$, then the disobedience planner again need only ensure that $dN > A$. This will ensure that the benign leaders will change policies, and it is impossible to get the tough leaders to change policies, since repression is always less costly. In this case, protest is beneficial for the planner as long as $(1 - p_0)Y > C$.

If C'(d)>0, then the disobedience planner would always choose the minimal level d needed to elicit bargaining. If $K > A > \varphi K$, then d=N/A, if protest occurs. If $\varphi K > A$, then the d will equal A.

In this case, the model predicts the disobedience will be more common when $p_i$ is low and leaders are more likely be benign. This fact suggests that historical theories that try to explain disobedience by focusing on bad government may be looking in the wrong direction. Tough dictators are less attractive targets for civil disobedience than benign rulers, who have higher costs of brutal suppression. In a sense, this simple point is no more than the algebra behind Tilly, Tilly and Tilly's (1975) empirical statement about rioting: "repression works." This fact would only be exacerbated if benign leaders also had lower costs of accommodation.

A notable difference between the archaic case and the modern case is that in the archaic case there is really no downside to extreme civil disobedience. The goal is to cause pain to the sovereign who can alleviate that pain by changing policies. Given that logic, perhaps it is unsurprising, that civil disobedience led to rebellion, for the logic of this non-informational protest is just to cause enough pain to induce the leader to change his policies.

## V. Bounded Rationality and Civil Disobedience

Our model presents a benchmark setting where civil disobedience is rational persuasion. The act of disobedience teaches voters about the preferences of the protesters or the nature of the leader. Rational protesters interested in the first objective will just choose an epsilon protest. The second objective calls for a more pain-producing protest that will elicit repression by the tough leaders but not the benign leaders.

Yet the behavior of protesters doesn't necessarily seem to be as rational as the model suggests. The whole field of civil disobedience is one in which rational calculation has rarely seemed at the forefront. We here discuss the ways in which behavior by the different actors may be less than fully rational in the sense of that they act in a way that is counter to their personal interests. In the next section, we discuss other psychological factors.

We have illustrated how disobedience can be a rational tool for toppling disliked leaders. Yet there are reasons why we might think that the conditions for such disobedience are unlikely to be met within the U.S. It seems quite possible that $\gamma < 0$, and voters prefer tough leaders, which seems to have been empirically true that American history post-1968. Moreover, it also seems quite likely than many disobedience planners have little private information about the nature of the leader.

If protest planners have little private information, then their protest intrinsically will do little to generate unhappiness with the current regime. Their one hope is that the protest will reveal information about the regime. In principle, the value of information is the same whether voters like toughness or mildness. In either case, a separating equilibrium causes the types to become known, which is desirable to the planner if his preference are convex.

Yet if protest planners would like to eliminate more repressive regimes, their strategies seem likely to have the opposite impact. While disobedients from a particular group signal their unhappiness by protesting. But their unhappiness does not do much to convince the median voter that the leadership is bad. On the contrary, a forceful response might convince the median voter that the leadership is good – which is a plausible account of the reaction of California voters to the tough response of Governor Ronald Reagan in the late 1960s.

We seem to observe significantly more civil disobedience than would be expected on purely rational grounds, at least if the purpose of disobedience is to produce reform. We occasionally see protest repression which seems far more extreme than would seem to optimal from the perspective of the leader.

We now consider whether bounded rationality among disobedients, voters and leaders can explain why civil disobedience appears to occur even when the conditions implied by our model are not met. In the case of the protesters, we will not need any additional mathematics. It is fairly obvious why errors by disobedients can engender more civil disobedience since the disobedients are making the disobedience decision themselves. In the case of voters and leaders, the impact on disobedience works through the behavioral choice of the disobedients. We will therefore use two illustrative models that illustrate the ways in which boundedly rational voting or leadership can lead to more civil disobedience.

*Bounded Rationality among the Disobedients*

We now discuss several additional behavioral complications that can explain why civil disobedience seems more common than our hyper-rational model might suggest. It is true that people might engage in civil disobedience even if the likelihood of triggering reform is low: To the disobedients, the costs of disobedience might be quite low compared to the discounted benefits, even if the realistic probability of success is under (say) 10 percent. Nonetheless, there are reasons to think that those who engage in disobedience will often exaggerate their prospects.

We differentiated between the behavior of the individual disobedients and the disobedience planners. The size of the protest, N, is a solution to the equation $1 - G\left(\pi c(N) - b_j\right) - \frac{N}{Q} = 0$. In this equation, bounded rationality might distort the perceived value of $\pi$ (the anticipated probability of a crack-down), $b_j$ (the taste for disobedience) or Q (the perceived size of the aggrieved minority). The disobedience planner is more like to plan a significant action if he perceives F(.) to be convex and if he expects the equilibrium number of disobedients to be significant. As such, if the planner erroneously overestimates the value of $b_j$ or Q are high, then disobedience will become more appealing.

Those reasons come from well-established behavioral findings, but we emphasize that their application here is speculative and requires further empirical work.

(1) *Unrealistic optimism*. In many contexts, human beings are prone to "optimistic bias" – to an unrealistically rosy sense of their future prospects (Sharot 2010). Optimistic bias can be seen as a form of motivated reasoning -- as, for example, where people believe that they are not subject to certain health risks because they are strongly motivated to have that belief. If people engage in civil disobedience with the goal of changing the status quo, there is a risk of excessive optimism about the actual prospects. In a sense, that optimism can be helpful insofar as it is motivating. But it might well lead people to engage in civil disobedience with an exaggerated sense of the probability of success.

In the context of our model, this might mean that disobedients believe that $\pi$ is too low, which would reduce the perceived costs of disobedience and make the size of disobedience larger. If the planner believes that the individual protesters underestimate the probably of repression than the planner will anticipate larger crowds and that makes protest more likely. Holding the beliefs of the individual diobedients constant, the planner himself is more likely to protest if he overestimates the value of $\pi$, since suppression is likely to make the protest more effective in generating regime change. Overoptimism could well lead the protest planner to have an overly high assessment of the probability of repression (since that is what he is hoping for) and for the individual disobedients to have an overly low assessment of the repression probability (since they are hoping not to experience repression).

Over-optimism could also function by changing the planners perceived assessment of the function F(.)—the voters preferences. An overly optimistic planner might be more likely to believe that voters will oust a leader who reveals himself to be tough and that will also make disobedience more attractive.

(2) *Overestimating one's representativeness # 1*—which basically means that if you believe that a leader is cruel or unjust, then others will share your dislike of him. Consider the robust psychological finding of "egocentric bias": People tend to think that other people share their values and tastes. Those who are engaged in civil disobedience are highly likely to think that other people are likely to do so, or at least share their opinions. Here as well, the result can be an inflation of the prospects for success.

In the model, this would be reflected by an overassessment of the value of Q, the size of the aggrieved group, by either the disobedients themselves or the disobedience planner. If the disobedients themselves make the error, then the protest will be larger because anticipating a bigger crowd means anticipating less downside if a crack-down does occur. If the planner makes the error, then this will make a protest more likely because they planner is more likely to think that he can engineer a really effective protest.

(3) *Overestimating one's representativeness # 2*—thinking that more people share your taste for civil disobedience. Egocentric bias might also lead people to believe that others will be willing to engage in disobedience. It should be unsurprising to find that protestors often offer greatly inflated predictions of the numbers of people involved (though the inflation might also be strategic).

This could also be captured by an exaggerated assessment of the size of Q by either the individual disobedients or the planner. This could also be captured by an overassessment of the value of $b_j$ by the planner, which will also lead him to anticipate a larger protest and then will make disobedience more appealing.

(4) *Outrage*. Those who engage in civil disobedience are often motivated in whole or in part by outrage: They believe that they are responding to serious injustice. When this is so, it is natural for people to think in retributive rather than strategic terms and to act as if the goal is to punish or to inflict pain on the authority, regardless of the effects of doing so (see Kahneman et al. 2000). If some or many of the disobedient are seeking to impose such punishment, they might fail to promote reform (and might not much care if they do fail). One result of outrage is to produce futile or counterproductive acts of disobedience.

In our model, outrage would reflect a larger value of $b_j$ which would make disobedience more attractive both to the individual disobedients and to the planner.

(5) *Group dynamics*. The psychological mechanisms do not, of course, operate in a social vacuum. They are greatly affected by group dynamics and social influences. A central phenomenon is group polarization (Glaeser and Sunstein 2009), by which, *members of a deliberating group usually end up adopting a more extreme version of the position toward which they tended before deliberation began*. The problem is especially severe for groups of like-minded people, who typically become more extreme as a result of deliberation (Sunstein and Hastie 2014). Group polarization has been found in hundreds of studies involving more than a dozen countries, including the United States, France, Afghanistan, and Germany. In the context of civil disobedience, group polarization is likely to play a serious role, and in multiple respects. If members of the group begin with certain substantive views, and if they talk and listen mostly to one another, those views will be heightened. The same is true is they suffer from unrealistic optimism or egocentric bias, or if they begin with high levels of outrage.

Here, then, is a psychological account of why civil disobedience will sometimes occur when a rational assessment, from the standpoint of group members, suggests that it should not: Social influences will interact with individual tendencies to produce action that is unlikely to produce reform. In practice, of course, it will not be easy to separate

inflated probability estimates from rational "moon shots" and from acts of conscience, which support disobedience whether or not it is likely to be successful. This structure would require a more complex formal model, but we hope that the basic logic is nonetheless clear.

*Boundedly Rational Voters*

Bounded rationality can also generate civil disobedience even if the disobedients themselves are completely ration. We first examine the possibility that voters are boundedly rational and examine what that does to the incentives to protest. We next examine the possibility of boundedly rational leadership.

The rational voter literature typically suggests that very act of voting is not particularly compatible with hyper-rational instrumental activity, but we don't mean to reenter that ancient debate. Our question is whether voters appear to respond to civil disobedience in ways that are compatible with rational inference. Did a protest change people's opinions by revealing something significant?

The quantitative version of this question is whether protests that revealed something meaningful effected more change in voter sentiment than protests that revealed little. Certainly, there are many protests that seem to have revealed little and that also did little change the political equilibrium, but it isn't obvious that the most successful protests, including the Civil Rights movement in the America South, revealed all that much pure information.

The facts about the American Jim Crow south had been presented by reasonable sources, notably Gunnar Myrdal, years before the Civil Rights movement became nationwide news. It is perhaps plausible that the Civil Rights movement made northerners more aware of the degree of unhappiness in the South. It is also plausible that the movement revealed the extreme unpleasant character of this repressive regime in a distinctly stark fashion.

Yet that revelation can hardly have been starker than the numbers of lynchings reported in the north before World War II that seem to have had little impact on northern attitudes. The preferences in the north surely shifted somewhat before and after the war, perhaps because of the experience of fighting a war against a virulently racist regime. Yet still, it is hard to think that the civil rights movement is best seen as a purely informational success. Similar arguments are easy to make about Gandhi's success with British popular opinion.

An alternative view is that information obtained through reading Myrdal's American Dilemna is just radically less salient that information obtained by watching the evening news and seeing peaceful protesters being attacked by hoses and dogs by Southern police officers. The salient, moving images are just vastly more effective at shaping public opinion than the printed word.

We can model this gap within the context of the model with a slight deviation from standard extreme rationality. We assume that citizens will not make any inference about the leaders' toughness from peaceful protest—and that as long as the leader is not actively repressing a protest, the citizens believe that the leader is tough with probability $p_0$. Moreover, we look at the behavior of protest planner who knows the leader is tough with probability $\sigma p_0$, which may be substantially higher than $p_0$. We assume that Q is sufficiently high and $\xi$ is sufficiently close to zero so that $\pi_{max} = 1$ and $\widehat{D} > \varphi K + V(F(\gamma - \theta_L) - F(-\theta_L))$:

*Proposition 7:* If voters do not learn from non-repressed protests, then if F(.) is concave on the interval $[-\theta_L, \gamma - \theta_L]$, there exists a value of $\sigma$ between 1 and $1/p_0$ at which protest planners are indifferent between doing nothing and setting d to generate a fully separating equilibrium. For higher values of $\sigma$, the planner's strictly prefer a separating equilibrium while for lower values of $\sigma$ the planner prefers doing nothing. The value of $\sigma$ which makes planners indifferent between doing nothing and protest is

Proposition 7 addresses a very simple form of limited rationality among voters that can readily generate disobedience and repression. The Proposition assumes that voters just won't be persuaded by any form of mass action. Such actions are just assumed to be non-salient and they don't move the opinons of the voters. Voters will only change their beliefs when they see repression.

We therefore consider the choice of planners' who choose between enough protest to generate a fully separating equilibrium where tough leaders repress and mild leaders do not and doing nothing, where voters persist in their old beliefs. The planner however has private information, and therefore he thinks that his protest is relatively more likely to create repression. Thus on average, the repression will occur more likely than the voters' expect it too.

This generates an incentive for significant protest even when the returns are concave. If the planners' beliefs diverge sharply from those of the voters then he will choose to risk repression. A failure to internalize non-salient information opens a gap in beliefs between protesters and voters if salient protest is the only way to change voters' beliefs then this will generate civil disobedience.

We do not think that this discussion diverges too strongly from the rational case discussed above. We have just emphasized that voters might need salient signals, not just solid information, which would be natural if there were costs of processing information. Voters have few sharp incentives to figure out political facts. A highly salient act of disobedience with particularly salient repression is just one means of reduce the costs of processing information.

According to this view, northern Americans were vaguely aware of the state of the American South in 1955 but they hadn't bothered to really figure out what that meant for the well-being of southern African-Americans or the brutality of the regime. The protests broadcast searing images that lowered the cost of information. If this is correct, then Martin Luther King didn't

have to face convex returns. He just had to hold a belief that the Southern regime was brutal which was higher than the belief of the average northerner.

Another interpretation of Proposition 7 is that it is the protest planner, rather than the general public, who is in error. In this case, the public prior on the leaders' type is correct, but the planner incorrectly believes that the leader is more likely to be repressive and the public is wrong. This error will also lead to more civil disobedience, but if the error is on the part of the planner then the disobedience is unlikely to be successful.

This is surely not the only plausible form of bounded rationality in the voting both, but it is a simple possibility and it illustrates one way in which bounded rationality among voters can make civil disobedience more appealing even if the disobedients are rational themselves.

*Boundedly Rational Leaders and Dyer's Error*

If voters are rational and disobedients are rational, then it is still possible for bounded rationality on the part of the leader to encourage civil disobedience. Since the disobedience planner and the leader have opposing objectives, the planner's incentives to protest typically increase when the leader is bad at serving his own interests.

Given that the behavior of leaders is based on unobservable tastes for repressing disorder, it would be impossible to ever claim that they acted irrationally. The narrower question is whether leaders act in ways that are contrary to their career concerns. Do we see political leaders regularly making decisions that come back to haunt them in their careers or at the voting booths?

The traditional "error", which we associate with Reginald Dyer is an overly tough response to mild disorder that plays havoc with a leader's career. Dyer was the Brigadier General in the Indian Army who gave the order to open on a crowd of thousands of peaceful Indians who were celebrating the Baisakhi festival. Martial law had been declared, so that the celebrants were violating the law, whether consciously or not. Dyer's response involved ordering over ten minutes of firing at the thick of a crowd that was trying to disperse causing the deaths of hundreds.

The Amritsar massacre would not be remarkable if perpetrated by an autocratic regime, but it is unusual by Dyer's superiors were ultimately responsible to British (but not Indian) voters. While there was certainly a vocal minority that strongly supported Dyer, his own career was over. The force of public response required the end of Dyer's career. British horror at British brutality in India has been often thought to represent a turning point in the history of the Raj.

Naturally, there are many ways of arguing that Dyer was fully rational. While his military career was ended, he did receive an enormous fund for his retirement raised privately raised by Rudyard Kipling and the Morning Post. He seems to have had a taste for brutal repression over

accommodation. Perhaps most reasonably, prior to World War I, British repression in India had received far less opprobrium and perhaps he just made a mistake, not recognizing how the sensitivity to slaughtering hundreds of unarmed civilians might have been heightened. Any action can be rational and still a mistake.

When leaders make Dyer's error, then the benefits of protesting to the planner (if not the individual protester) can increase. Since the disobedience planner anticipates this overly heavy response, the planner knows that the protest is more likely to be revealing. This in turn makes protest more appealing.

One can plausibly argue that Southern policemen who turned the hoses and dogs on civil rights protesters were guilty of their own version of Dyer's mistake, because their actions helped end the regime they were allegedly trying to support. Yet from a private career's concern model, it is far less obvious that a Southern sheriff had much to fear from getting tough on protesters. The voters of the south eagerly returned tough leaders at least until African-Americans made up a large fraction of the voting public. A reasonable view is that each individual sheriff who was repressive acted in their own private career's interest, but that taken as a whole, the Southern leaders might have done better (strictly for their own long term survival) by agreeing upon more accommodation.

Two particularly salient examples of leaders who appeared to suffer because of a tough response to protest are Richard J. Daley and New York Police Chief Kelly. Daley's public image suffered significantly in response to the Chicago Police's tough tactics when faced with the protests at the 1968 Democratic Convention. Yet Daley was re-elected by his Chicago voters until his death. The New York Police Department's handling of Occupy Wall Street has been seen as helping to support the progressive anti-Bloomberg wave that helped elect DeBlasio and replace Kelly.

These individual cases may represent repression with negative longer term consequences, but they are rare. Since the 1960s, police response to disorder within the U.S. has generally been muted and the public has rarely punished any leader for heavy-handedness. That track record leads us back to where this section began—the protest planners themselves seem most likely to be influenced by bounded rationality.

## VI. Prosecutorial Discretion and Civil Disobedience

Our emphasis has been on the use of force by authorities, with the canonical example being the invocation of police. But the same analysis applies to a more common dilemma, by which prosecutors must decide whether to initiate proceedings against people who have engaged in various forms of civil disobedience. Consider, for example, tax protestors, who refuses to file because of disapproval of actions of the national government; civil rights protestors who have

unlawfully occupied streets and refused to disperse; or journalists who refuse to divulge their sources (in violation of the law, and on principle). In cases of this kind, should prosecutors bring the force of law to bear, or should they exercise their discretion so as to leave the actions unpunished?

There are of course moral questions about the appropriate approach (Dworkin 1967), but prosecutors should also ask a purely instrumental question, which is whether prosecution will be helpful or harmful if underlying goal is to prevent repetition of the underlying behavior. In the abstract, it could go either way. Like any other use of force, prosecution might increase the resolve of those who are agree with the law-breaker (and promote imitation), or it might serve instead to deter them.

The work in this paper highlights that the power to persuade will lie in the response far more than in the initial action. While disobedients seem likely to have little information relative to the general public, the authorities surely know themselves. As such, their actions always have a great deal of power to inform the public. This suggests an asymmetry between the protesters, whose ability to screw-up is minimal, and the police, whose ability to screw up is enormous.

In a rational model, there are two primary reasons why disobedients think that rolling the dice and potentially eliciting a response may be sensible. First, it may be that the returns are convex, and in our model, this came from a convexity in the distribution of voter tastes. But more generally, the requirement is that revealing a leader to be tough is far more important than allowing a leader to remain appearing innocuous.

Disobedience will also become more attractive to individuals who believe that they know that the leaders are truly bad guys. They think that the leader will respond harshly with a high probability and hence they want to provoke the leader. Notably, this would be true if their beliefs were rational or if they had a less accurate but still negative view of their leaders.

It seems quite possible that a degree of bounded rationality may help explain the popularity of disobedience given its poor track record of creating change in recent decades. As the model has emphasized, individual protesters must themselves be moved by intrinsic motivation. This is the civil disobedience equivalent of the voters' paradox. Since no individual disobedient makes a difference, they must like doing it.

Nonetheless, in our model, the planning of the protest required instrumental aims, and we are left wondering whether the conditions needed for rational disobedience planning are actually satisfied in the U.S. in recent decades. Harsh responses occasionally occur, but most American voters seem comfortable re-electing tough leaders. Indeed, it seems quite possible that most U.S. leaders are more concerned with acting tough than pretending to be easy-going. In that case, it is quite difficult to generate rational disobedience. If this is the case, the disobedience is more likely to reflect either over-optimism or intrinsic motivation than a rational plan to inform voters.

## VII. Conclusion

Civil disobedience is often expressive rather than instrumental; people sometimes violate the law because of the perceived dictates of conscience. In many of the most interesting cases, and those that are most important from the standpoint of history, disobedience is designed to produce reform, either by changing public opinion (by showing the number and intensity of opponents to practices or regimes) or by triggering a forceful response that seems to reveal that the regime has a bad character.

Disobedience is a strong form of protest, because those who engage in it are risking sanctions. For that reason, the signal of their actions may become very loud. In some occasions, they can make misconduct or injustice salient when it would otherwise be in the background. On other occasions, the signal will convey information about the depth and width of public disapproval (Lohmann 1994). At the same time, the signal might alter reputational incentives, in part of dissipating pluralistic ignorance, showing people that opposition is far more widespread than they previously thought (Kuran 1998).

Much of our focus here has been on a particular motivation for civil disobedience, which is to provoke the leader to respond in ways that will shift public opinion in significant ways. For the disobedient, the challenge is to avoid action that is so weak that it fails to provoke or so aggressive that even the most forceful response will seem justified. We have explored the circumstances in which rational disobedience can surmount that challenge, consistent with some of history's most celebrated examples. At the same time, we have emphasized that optimism bias, egocentric bias, and outrage can produce futile or counterproductive disobedience, consistent with what is commonly observed.

**References**


Ashenfelter, Orley, and George E. Johnson. "Bargaining theory, trade unions, and industrial strike activity." *The American Economic Review* (1969): 35-49.

Banks, Jeffrey S., and Joel Sobel. "Equilibrium selection in signaling games." *Econometrica: Journal of the Econometric Society* (1987): 647-661.

Cho, In-Koo, and David M. Kreps. "Signaling games and stable equilibria." *The Quarterly Journal of Economics* (1987): 179-221.

DiPasquale, Denise, and Edward L. Glaeser. "The Los Angeles riot and the economics of urban unrest." *Journal of Urban Economics* 43, no. 1 (1998): 52-78.

Farber, Henry S. "Bargaining theory, wage outcomes, and the occurrence of strikes: an econometric analysis." *The American Economic Review* (1978): 262-271.

Fernández, Raquel, and Jacob Glazer. "Striking for a Bargain between Two Completely Informed Agents." *American Economic Review* 81, no. 1 (1991): 240.

Glaeser, Edward L. "The political economy of hatred." *The Quarterly Journal of Economics* (2005): 45-86.

Glaeser, Edward L., and Cass R. Sunstein. "Extremism and Social Learning." *Journal of Legal Analysis* 1, no. 1 (2009): 263-324.

Granovetter, Mark. "Threshold models of collective behavior." *American journal of sociology* (1978): 1420-1443.Jun (1989)

Kuran, Timur. "Ethnic norms and their transformation through reputational cascades." *The Journal of Legal Studies* 27, no. S2 (1998): 623-659.

Lohmann, Susanne. "The dynamics of informational cascades: the Monday demonstrations in Leipzig, East Germany, 1989–91." *World politics* 47, no. 01 (1994): 42-101.

Murray, Charles. "By the People: Rebuilding Liberty Without Permission." Crown Forum, 2015.

Sunstein, Cass R., and Reid Hastie. *Wiser: Getting beyond groupthink to make groups smarter*. Harvard Business Press, 2014.

Tilly, Charles, Louise Tilly, and Richard Tilly. *The rebellious century: 1830-1930*. Cambridge, MA: Harvard University Press, 1975.

**Table 1: A Typology of Disobedience**

| | Non-Instrumental | Instrumental |
|---|---|---|
| Violent | Disorganized Riot | Storming the Bastille<br><br>Wat Tyler's Rebellion<br><br>Boston Tea Party |
| Non-Violent Large Group | Occupy Movement | Gandhi's Salt March<br><br>Colonial Boycotts of British Goods<br><br>1926 U.K. General Strike |
| Non-Violent Small Group | | Freedom Rides<br><br>Ferguson Disobedience |
| Non-Violent Individual | Thoreau | Mohamed Bouazizi |

I.

**Table 2: The Response to Civil Disobedience**

| Leader Response | Minimum Value of V | Maximum Value of V |
|---|---|---|
| | $K > D > \varphi K$ | |
| Both types tolerate | $\frac{D - \varphi K}{F(\gamma - \theta_L) - F(\gamma p_D - \theta_L)}$ | None |
| Some Malign Suppress; All Benign Tolerate | $\frac{D - \varphi K}{F(\gamma - \theta_L) - F(-\theta_L)}$ | $\frac{D - \varphi K}{F(\gamma - \theta_L) - F(\gamma p_D - \theta_L)}$ |
| All Malign Suppress; All Benign Tolerate | None | $\frac{D - \varphi K}{F(\gamma - \theta_L) - F(-\theta_L)}$ |
| | $D > K > \varphi K$ | |
| All leaders tolerate | $\frac{D - \varphi K}{F(\gamma - \theta_L) - F(\gamma p_D - \theta_L)}$ | None |
| Some Malign Suppress; All Benign Tolerate | $\frac{D - \varphi K}{F(\gamma - \theta_L) - F(-\theta_L)}$ | $\frac{D - \varphi K}{F(\gamma - \theta_L) - F(\gamma p_D - \theta_L)}$ |
| Malign leaders suppress; benign leaders tolerate | $\frac{D - K}{F(\gamma - \theta_L) - F(-\theta_L)}$ | $\frac{D - \varphi K}{F(\gamma - \theta_L) - F(-\theta_L)}$ |
| All Malign $(1 - \varphi)K < F(\gamma - \theta_L) - F(\gamma p_0 - \theta_L)$, Suppress; Some Benign Tolerate | $\frac{D - K}{F(\gamma - \theta_L) - F(-\theta_L)}$ | $\frac{D - K}{F(\gamma p_D - \theta_L) - F(-\theta_L)}$ |
| All leaders Suppress | None | $\frac{D - K}{F(\gamma p_D - \theta_L) - F(-\theta_L)}$ |

**Appendix I: Proofs of Propositions**

Proof of Proposition 1: A useful lemma for all subsequent proofs is that it is impossible for both types of leaders to randomize between the two strategies. We let $p_{NS}$ reflect the posterior belief that the leader is malign conditional upon the leader not suppressing the disorder and $p_S$ reflect the posterior belief that the leader is malign conditional upon the leader suppressing the disorder. For the malign leader to be indifferent between suppressing and not suppressing disorder, it must be that $(1 - F(\gamma p_{NS} - \theta_L))V - D = (1 - F(\gamma p_S - \theta_L))V - \varphi K$. If this condition holds, then $(1 - F(\gamma p_{NS} - \theta_L))V - D > (1 - F(\gamma p_S - \theta_L))V - K$ and the benign leader will always strictly prefer non-suppression. As such, at least one type of leader will always follow a single strategy with probability one.

The same logic implies that if malign leaders choose not to suppress with any positive probability, which implies that $(1 - F(\gamma p_{NS} - \theta_L))V - D \geq (1 - F(\gamma p_S - \theta_L))V - \varphi K$, then benign leaders will never suppress because $(1 - F(\gamma p_{NS} - \theta_L))V - D > (1 - F(\gamma p_S - \theta_L))V - \varphi K$. Similarly, if benign leaders suppress with any positive probability, which implies that $(1 - F(\gamma p_{NS} - \theta_L))V - D \leq (1 - F(\gamma p_S - \theta_L))V - K$, then malign leaders will always suppress because $(1 - F(\gamma p_{NS} - \theta_L))V - D < (1 - F(\gamma p_S - \theta_L))V - \varphi K$. This implies that the only possible semi-pooling equilibria are ones in which all benign leaders and some malign leaders choose not to suppress or ones in which all malign leaders and some benign leaders suppress harshly. As such, it follows that in all semi-pooling equilibria, $p_S \geq p_{NS}$, voters always believe that harsh suppression is more likely to reflect a malign leader.

If $K > \varphi K > D$, then the cost of disorder is less than the cost of suppression for both types of leaders. Consider an equilibrium in which both types of leaders do not suppress disorder. If either type deviates to suppressing disorder, they will receive an immediate welfare loss, because the costs of suppression are greater than the costs of disorder, and they will be thought to be malign, which will reduce their chances of re-election. As such, this is an equilibrium.

Consider a candidate pooling equilibrium in which both types of leaders harshly suppress disorder. If either type deviates to non-suppression, they will receive an immediate welfare gain, because the costs of suppression are greater than the costs of disorder, and they will be more likely to be re-elected by voters will perceive them as being benign. As such, this cannot be an equilibrium.

No semi-pooling equilibrium is possible if $K > \varphi K > D$, because leaders will always lose by suppressing harshly. They lose directly because the costs of suppression are higher than the costs of disorder and they lose indirectly because the voters will be more likely to think that they are malign. As a result, they will always prefer non-suppression.

We now adopt the notation: $\vartheta_1 = 1 - F(\gamma - \theta_L)$, $\vartheta_0 = 1 - F(-\theta_L)$ and $\vartheta_D = 1 - F(\gamma p_D - \theta_L)$, where $\vartheta_1 < \vartheta_D < \vartheta_0$. $\vartheta_1$ reflects the probability of winning re-election if voters believe

that the leader is malign with probability one. $\vartheta_0$ reflects the probability of winning re-election if voters believe that you are malign with probability zero. $\vartheta_D$ reflects the probability of winning re-election if voters believe that you are malign with probability $p_D$, which will be true in any pooling equilibrium.

We first consider the case where $\gamma > 0$. A separating equilibrium, where all malign leaders suppress harshly and all benign leaders do not suppress, is possible if and only if $\vartheta_1 V - \varphi K \geq \vartheta_0 V - D \geq \vartheta_1 V - K$. These conditions ensure that neither type will want to deviate from the pooling equilibrium. This condition can be broken into two conditions on V, that $\frac{D-\varphi K}{\vartheta_0-\vartheta_1} \geq V \geq \frac{D-K}{\vartheta_0-\vartheta_1}$. If $K > D > \varphi K$, then the separating equilibrium exists when V is low (i.e. below $\frac{D-\varphi K}{\vartheta_0-\vartheta_1}$). If $D > K > \varphi K$, then the separating equilibrium can only exist if V is neither too low nor too high, i.e. between the two bounds.

A pooling equilibrium where both types suppress harshly can exist if and only if $\vartheta_D V - K \geq \vartheta_0 V - D$. This condition ensures that the benign leaders will not want to deviate, which ensures that the malign leaders will also not want to t deviate. This condition requires that $\frac{D-K}{\vartheta_0-\vartheta_D} \geq V$. As such, this equilibrium can only exist if $D > K > \varphi K$, and if V is sufficiently low. Notably $\frac{D-K}{\vartheta_0-\vartheta_1} < \frac{D-K}{\vartheta_0-\vartheta_D}$, so there exist values of V for which both this pooling equilibrium and the separating equilibrium both exist, as long as $D > K$.

A pooling equilibrium where both types do not suppress can only exist if $\vartheta_D V - D \geq \vartheta_1 V - \varphi K$, which ensures that the malign types will not want to deviate, which implies that the benign types will also not want to deviate. This condition requires that $\frac{D-\varphi K}{\vartheta_D-\vartheta_1} \leq V$. If $D < \varphi K$, this would always hold, which is why this equilibrium always exists in that case. If $D > \varphi K$, then this equilibrium only exists if V is sufficiently high to compensate the malign leaders with sufficient career returns to offset the immediate loss in utility.

As $\frac{D-\varphi K}{\vartheta_D-\vartheta_1} \geq \frac{D-\varphi K}{\vartheta_0-\vartheta_1}$, then minimum cutoff needed for this pooling equilibrium is above the maximum cutoff needed for the existence of separating equilibrium.

There is one semi-pooling equilibrium in which malign types randomize between the two actions, while the benign types always fail to suppress. We denote the probability of re-election condition upon semi-pooling at non-suppression as $\vartheta_{NS}$, which will lie between $\vartheta_0$ and $\vartheta_D$. The indifference condition needed for semi-pooling is that $\vartheta_1 V - \varphi K = \vartheta_{NS} V - D$ or $\vartheta_{NS} = \vartheta_1 + \frac{D-\varphi K}{V}$. The range of values for $\vartheta_{NS}$, then imply that this semi-pooling equilibrium can only occur if $\frac{D-\varphi K}{\vartheta_D-\vartheta_1} \geq V \geq \frac{D-\varphi K}{\vartheta_0-\vartheta_1}$. If the V values are too high, then the malign types will always want to imitate the benign types and not suppress. If the V values are too low, then the malign types will never want to do anything other than punish harshly, as long as D>$\varphi$K.

There is a second semi-pooling equilibrium in which benign types randomize between the two actions, while the malign types always suppress harshly. In this case, we denote the probability of reelection conditional upon semi-pooling at harsh suppression as $\vartheta_S$, which will fall between $\vartheta_1$ and $\vartheta_D$. The indifference condition needed for semi-pooling is that $\vartheta_S V - K = \vartheta_0 V - D$, or $\vartheta_S = \vartheta_0 - \frac{D-K}{V}$. Obviously, a necessity for this equilibrium to exist is that D>K, so that benign types lose in current welfare from non-suppression. The range of values of V for which this equilibrium exists is that $\frac{D-K}{\vartheta_0-\vartheta_D} \geq V \geq \frac{D-K}{\vartheta_0-\vartheta_1}$.

Using these conditions, we can characterize the equilibria that can exist under the two remaining configurations for D. If $K > D > \varphi K$, then there is no semi-pooling or pooling equilibrium where the benign suppress harshly. The benign prefer not suppressing and voters will always be more likely to think that they are benign if they don't suppress. There is a semi-pooling equilibrium where the malign do not suppress as long as $\frac{D-\varphi K}{\vartheta_D-\vartheta_1} \geq V \geq \frac{D-\varphi K}{\vartheta_0-\vartheta_1}$. There is a separating equilibrium when $\frac{D-\varphi K}{\vartheta_0-\vartheta_1} \geq V$. There is a pooling equilibrium where both types do not suppress when $\frac{D-\varphi K}{\vartheta_D-\vartheta_1} \leq V$. As such, generically, there is a unique equilibrium in this cases, and the equilibrium is determined by the value of V. For high values of V, all types pool on leniency. For low values of V, there is separation. For intermediate values of V, there is pooling, where some, but not all, of the malign types imitate the benign types with leniency.

If $D > K > \varphi K$, then there is a pooling equilibrium where both types suppress exists if $\frac{D-K}{\vartheta_0-\vartheta_D} \geq V$. A semi-pooling equilibrium in which benign types randomize between the two actions, while the malign types always suppress harshly exists if $\frac{D-K}{\vartheta_0-\vartheta_D} \geq V \geq \frac{D-K}{\vartheta_0-\vartheta_1}$. A pooling equilibrium where both types don't suppress exists if $\frac{D-\varphi K}{\vartheta_D-\vartheta_1} \leq V$. A semi-pooling equilibrium in which malign types randomize between the two actions, while the benign types always fail to suppress if $\frac{D-\varphi K}{\vartheta_D-\vartheta_1} \geq V \geq \frac{D-\varphi K}{\vartheta_0-\vartheta_1}$. A separating equilibrium exists if $\frac{D-\varphi K}{\vartheta_0-\vartheta_1} \geq V \geq \frac{D-K}{\vartheta_0-\vartheta_1}$.

If $V < \frac{D-K}{\vartheta_0-\vartheta_1}$, the unique equilibrium is pooling where both types suppress. If $\frac{D-K}{\vartheta_0-\vartheta_D} > V > \frac{D-K}{\vartheta_0-\vartheta_1}$, then the generically there always exists three equilibria. There is always a pooling equilibrium in which both sides suppress harshly. There is also semi-pooling equilibrium in which benign types randomize between the two actions, while the malign types always suppress harshly. Finally, there is a third equilibrium, which his either separating (if $\frac{D-\varphi K}{\vartheta_0-\vartheta_1} \geq V$), semi-pooling if $\frac{D-\varphi K}{\vartheta_D-\vartheta_1} \geq V \geq \frac{D-\varphi K}{\vartheta_0-\vartheta_1}$) or pooling where both types don't suppress if $\frac{D-\varphi K}{\vartheta_D-\vartheta_1} \leq V$. If $V > \frac{D-K}{\vartheta_0-\vartheta_D}$, then there is a unique equilibrium, which is either sepa rating (if $\frac{D-\varphi K}{\vartheta_0-\vartheta_1} \geq V$), semi-pooling if $\frac{D-\varphi K}{\vartheta_D-\vartheta_1} \geq V \geq \frac{D-\varphi K}{\vartheta_0-\vartheta_1}$) or pooling where both types don't suppress if $\frac{D-\varphi K}{\vartheta_D-\vartheta_1} \leq V$.

We now turn to the two cases considered when $\gamma < 0$ and all leaders would like to appear tough. In this case, if $D > K > \varphi K$, then it would be impossible for either type of leader to do nothing, since both get negative direct utility from doing nothing, and neither can benefit reputationally by doing nothing. As such both types of leaders will be tough.

IF $K > D > \varphi K$, then it will never be the case that tough types do nothing, since their reputation can only be hurt. Thus we must only consider the behavior of the benign types. For low enough values of V, it must be that the direct utility from action dominates the reputational consequences and a separating equilibrium exists. This equilibrium will exist as long as $(F(-\theta_L) - F(\gamma - \theta_L))V$ (the reputational benefit for benign type imitating tough types) is less than $K - D$ (the cost of imitating tough types). For high enough values of V, there is an equilibrium in which all of the benign types imitate the tough types, this equilibrium can only exist as long as $(F(-\theta_L) - F(\gamma p_D - \theta_L))V$ (the loss from deviating to doing nothing) is greater than $K - D$. There can also exist an equilibria in which the benign types mix between the two actions and that equilibrium requires that $\big(F(-\theta_L) - F(\gamma p_{Mix} - \theta_L)\big)V = K - D$, as $p_{Mix}$, the share of tough types conditional upon repressing can only range from $p_D$ to 1, the values of V for which this semi-pooling equilibrium exists ranges from $\frac{K-D}{F(-\theta_L)-F(\gamma-\theta_L)}$ to $\frac{K-D}{F(-\theta_L)-F(\gamma p_D-\theta_L)}$. As such, there is a unique equilibrium for every value of V.

*Proof of Proposition 2:* The separating equilibrium exists when $\frac{D-\varphi K}{F(\gamma-\theta_L)-F(-\theta_L)} \geq V$. This upper limit is increasing in $D$, and decreasing in $\varphi,\ K, \gamma$. Differentiation also yields that the limit is increasing with $\theta_L$ if and only if $f(-\theta_L) < f(\gamma - \theta_L)$. The pooling equilibrium where both types do not suppress occurs when $\frac{D-\varphi K}{F(\gamma-\theta_L)-F(\gamma p_D-\theta_L),} \leq V$. This lower bound is increasing in $D$ and $p_D$ and decreasing with $\varphi$, and K. The lower bound if increasing with $\theta_L$ if and only if $f(\gamma p_D - \theta_L) < f(\gamma - \theta_L)$. The lower bound is decreasing with $\gamma$ if and only if $p_D f(\gamma p_D - \theta_L) < f(\gamma - \theta_L)$.

$$F\left(\frac{\gamma x p_D}{1 - p_D + x p_D} - \theta_L\right)$$

In the semi-pooling equilibrium, the tough randomize between suppressing harshly and imitating the benign and being tolerant. We lot x denote the share of the malign who imitate, and this share must satisfy: $\left(1 - F\left(\frac{\gamma x(1-p_D)}{1-p_D+x p_D} - \theta_L\right)\right)V - D = (1 - F(\gamma - \theta_L))V - \varphi K$, or $F(\gamma - \theta_L) - \frac{D-\varphi K}{V} = F\left(\frac{\gamma x(1-p_D)}{1-p_D+x p_D} - \theta_L\right)$. This equality holds for some value of x between zero and one, which adjusts to make the equality hold. The left hand side of the equality is increasing with x, and hence x must be falling with D and rising with $\varphi$, K and V.

*Proof of Proposition 3:* Define z as the share of potential protesters who do protest, and v(z) as the value of $b_j + \varepsilon_m$ for the marginal protestor. We note that $v(z)$ satisfies $z = 1 - G\big(v(z) - b_j\big)$ and hence $\frac{1}{-g\left(v(z)-b_j\right)} = v'(z)$ and $-\frac{g\prime\left(v(z)-b_j\right)}{g\left(v(z)-b_j\right)^3} = v"(z)$. The first derivative is always negative, and the second derivative has the opposite sign of $g'\big(v(z) - b_j\big)$. The assumption that the distribution of $\varepsilon_m$ is single peaked at the median implies that $v"(z) < 0$ if z is less than ½.

The net benefit for the marginal disobedient is $v(z) - \pi c(zQ)$. Equilibria occur at the point in which $v(z)$ crosses $\pi c(zQ)$. The first derivative of $v(z) - \pi c(zQ)$, with respect to z is $\frac{1}{-g\left(v(z)-b_j\right)} - \pi Q c'(zQ)$, which is ambiguous in sign since c'(zQ)<0. The second derivative is $-\frac{g\prime\left(v(z)-b_j\right)}{g\left(v(z)-b_j\right)^3} - \pi Q^2 c"(zQ)$, which must be negative for z<.5.

For all values of $\pi$, $v(.5) - \pi c(.5Q) < 0$, as we have assumed that $v(0) > 0 > v(.5)$. Since v(z) is continuously decreasing, there must exist a unique value of z<.5 such that $v(z) - \pi c(zQ)$=0, if $v(0) > \pi c(0)$ or $\frac{\varepsilon_{max}+b_j}{c(0)} > \pi$. The equilibrium level of disobedience is rising with Q and falling with $\pi$, as $v(z) - \pi c(zQ)$ crosses zero only from above when $v(0) > \pi c(0)$.

If $\frac{\varepsilon_{max}+b_j}{c(0)} < \pi$, then one equilibrium will always involve no disobedience, but there may be others.

Let $w(z, Q)$, equal $v(z)/c(zQ)$, and let W(Q) equal the maximized value of w(z,Q), which is concave. If the derivative of w(z,Q) at zero is negative, or $\frac{1}{Q} > -g(\varepsilon_{max})\big(\varepsilon_{max} + b_j\big)\frac{c\prime(0)}{c(0)}$, then W(Q) will equal $\frac{\varepsilon_{max}+b_j}{c(0)}$, which is independent of Q. If $\frac{1}{Q} < -g(\varepsilon_{max})\big(\varepsilon_{max} + b_j\big)\frac{c\prime(0)}{c(0)}$, then w(z,Q) will be greater than zero and W(Q) will be greater than $\frac{\varepsilon_{max}+b_j}{c(0)}$. The value of W(Q) is monotonically increasing with Q and will eventually be greater than one.

If $\pi > W(Q)$, then for all values of z, $\pi c(zQ) > v(z)$ and no equilibrium can exist with positive disobedience. If $v(0)/c(0) < \pi \leq W(Q)$, then there must exist no just one but two values of z for which $\pi c(zQ) = v(z)$ as $v(z)/c(zQ)$ is concave and goes to zero as z goes to one-half and to $v(0)/c(0)$ as z goes to zero. Hence there must exist two crossing points that correspond to two equilibria. The third equilibrium has z=0. The level of z in the higher equilibrium which occurs when $\pi = v(z)/c(zQ)$ and $v(z)/c(zQ)$ is downward sloping with z and hence is increasing with Q and $b_j$ and decreasing with $\pi$.

The value of W(Q) is increasing with Q and the associated z must satisfy $\frac{v\prime(z)}{v(z)} = \frac{c\prime(zQ)Q}{c(zQ)}$, or $1 = \frac{-c'((1-G(\varepsilon^*)Q))Q}{c((1-G(\varepsilon^*)Q)}\big(\varepsilon^* + b_j\big)g(\varepsilon^*)$, where $\varepsilon^*$is the associated maximizing preference of the

marginal disobedient. The value of W(Q) or $\frac{\varepsilon^* + b_j}{c((1-G(\varepsilon^*)Q)}$ defines the highest possible value of $\pi$ for which an equilibrium with positive disobedience can occur. The maximum $\pi$ equals term $\frac{\varepsilon_{max}+b_j}{c(0)}$ if $\frac{1}{Q} > -g(\varepsilon_{max})(\varepsilon_{max} + b_j)\frac{c\prime(0)}{c(0)}$; if $\frac{1}{Q} < -g(\varepsilon_{max})(\varepsilon_{max} + b_j)\frac{c'(0)}{c(0)}$, for higher values of Q, the maximum $\pi$ rises montonically with Q reaching a maximum value of one.

The maximum value of $\pi$ is also increasing with $b_j$.

Proof of Lemma 1: There is exactly the same information revelation with the smallest value of d, as there is with any higher level of d if, in equilibrium, all leaders do the same thing. As such, the political benefits to the disobedience planner are the same, and we have assumed that the planner has a slight, but still real, preference for protests with the lowest possible level of d.

Proof of Proposition 4: The welfare to the planner from a pooling equilibrium, whether or not the leaders repress or tolerate, equals $B_H F(\gamma\sigma p_0 - \theta_L)$. Any separating equilibrium in which leaders take different actions will yield welfare of $B_H\big(\pi F(\gamma p_1 - \theta_L) + (1-\pi)F(\gamma p_2 - \theta_L)\big)$, where $\pi$ represents the unconditional probability that the leaders will take the first action, and $p_1$ represents the conditional expectation that the leader is tough if the leader takes action 1 (which we always assume is weakly greater than $p_2$, since we can define action 1 as the action taken more often by the tough leaders). Adding up requires that $\pi p_1 + (1-\pi)p_2 = \sigma p_0$, so substituting in yields that the planner's welfare equals

$B_H\left(\pi F(\gamma p_1 - \theta_L) + (1-\pi)F\left(\gamma\frac{\sigma p_0 - \pi p_1}{(1-\pi)} - \theta_L\right)\right)$. The derivative of this with respect to $p_1$ equals $\pi\gamma\left(f(\gamma p_1 - \theta_L) - f\left(\gamma\frac{\sigma p_0 - \pi p_1}{(1-\pi)} - \theta_L\right)\right)$, which is always negative if F(.) is concave in this region. Hence is always better for the planner to have the same share of tough and mild leaders taking each action. If the shares are the same, then there is no advantage in not having all leaders take the same action and hence the epsilon equilibrium is the best outcome possible.

Consider the equilibrium range in which $D \le \widehat{D}$, in which there is pooling with no repression for $D < \varphi K + V(F(\gamma - \theta_L) - F(\gamma\sigma p_0 - \theta_L))$ and semi-pooling with some repression if $\widehat{D} \ge D > \varphi K + V(F(\gamma - \theta_L) - F(\gamma\sigma p_0 - \theta_L))$. The planners welfare in this region equals $B_H$ times $(1-\mu\,(\mathrm{D}))\sigma p_0 F(\gamma - \theta_L) + (1-(1-\,\mu\,(\mathrm{D}))\sigma p_0)F\left(\frac{\gamma\mu(D)\,\sigma p_0}{1-(1-\,\mu(D)\,)\sigma p_0} - \theta_L\right)$, where $\mu$ (D) satisfies $D - \varphi K + VF\left(\frac{\gamma\mu(D)\sigma p_0}{1-(1-\,\mu(D))\sigma p_0} - \theta_L\right) = VF(\gamma - \theta_L)$, so $\mu'(\mathrm{D}) = -\frac{(1-(1-\,\mu(D))\sigma p_0)^2}{\gamma\sigma p_0(1-\sigma p_0)Vf\left(\frac{\gamma\mu(D)\sigma p_0}{1-(1-\,\mu(D))\sigma p_0} - \theta_L\right)}$. The derivative of this with respect to D equals $\frac{B_H(1-(1-\,\mu\,(\mathrm{D}))\sigma p_0)}{V}$ times $\left(F(\gamma - \theta_L) - F\left(\frac{\gamma\mu(D)\,\sigma p_0}{1-(1-\,\mu(D)\,)\sigma p_0} - \theta_L\right)\right)\frac{1-(1-\,\mu(D))\sigma p_0}{\gamma(1-\sigma p_0)f\left(\frac{\gamma\mu(D)\sigma p_0}{1-(1-\,\mu(D))\sigma p_0} - \theta_L\right)} - 1$, which is always positive if F(.) is convex over the region and negative if F(.) is concave. Hence

the planner will always choose the highest value of D, up to the point where D equals $\widehat{D}$ or full separation occurs which requires that $D = \varphi K + V(F(\gamma - \theta_L) - F(-\theta_L))$. If it is impossible to generate any repression, then the epsilon protest generates the best outcome possible.

Proof of Proposition 5: If $\pi_{max} > \sigma p_0$, then the planners can induce a protest even in his preferred fully separating equilibrium with repression from the tough leaders. As a result, the preferences of his followers do not restrict his actions and the results of Proposition 4 apply. If $\pi_{max} < \sigma p_0$, then he will not be able to achieve a fully separating equilibrium. The planner will face two constraints. D must be less than or equal to $\widehat{D}$ and the maximum value of $\mu(D)$ then will satisfy$(1 - \mu\ (\mathrm{D}))\sigma p_0 = \pi_{max}$, which implies $D - \varphi K + VF\left(\frac{\gamma(\sigma p_0 - \pi_{max})}{1 - \pi_{max}} - \theta_L\right) = VF(\gamma - \theta_L)$ The planner will prefer the highest level of $D$ (which is also the highest value of $(1 - \mu\ (\mathrm{D}))\sigma p_0$) to all lower levels, because of his convex preferences.

If $\pi_{max} < \sigma p_0$, then if $\varphi K + V\big(F(\gamma - \theta_L) - F(\gamma\sigma p_0 - \theta_L)\big) > \widehat{D}$, then again, no separation is possible and the epsilon protest dominates. If $\varphi K - VF\left(\frac{\gamma \pi_{max}}{1 - \sigma p_0 + \pi_{max}} - \theta_L\right) + VF(\gamma - \theta_L) > \widehat{D} > K + V\big(F(\gamma - \theta_L) - F(\gamma\sigma p_0 - \theta_L)\big)$, the planner will set d so that D equals $\widehat{D}$ and if $\varphi K - VF\left(\frac{\gamma \pi_{max}}{1 - \sigma p_0 + \pi_{max}} - \theta_L\right) + VF(\gamma - \theta_L) < \widehat{D}$, then the planner will choose d so that $d(1 - G(\varepsilon^*))Q - \varphi K + VF\left(\frac{\gamma(\sigma p_0 - \pi_{max})}{1 - \pi_{max}} - \theta_L\right) = VF(\gamma - \theta_L)$ where $\pi_{max} = \frac{\varepsilon^* + b_j}{c((1 - G(\varepsilon^*)Q)}$ and $\varepsilon^*$satisfies $c((1 - G(\varepsilon^*)Q) = -c'((1 - G(\varepsilon^*)Q)Q\big(\varepsilon^* + b_j\big)g(\varepsilon^*)$.

*Proof of Proposition 6:* If the planners is unconstrained and F(.) is convex, then $D = \varphi K + V(F(\gamma - \theta_L) - F(-\theta_L))$. In this case, D is rising with $\varphi$, K, V, $\gamma$ and falling with $\theta_L$ (as $f(\gamma - \theta_L) > f(-\theta_L)$ from the convexity of F(.) in this region.

If $\widehat{D} = D$, then $d\big(1 - G(\hat{\varepsilon})\big)Q = \widehat{D}$, $\widehat{D} - \varphi K + VF\left(\frac{\gamma \mu(\widehat{D})\sigma p_0}{1 - \big(1 - \mu(\widehat{D})\big)\sigma p_0} - \theta_L\right) = VF(\gamma - \theta_L)$, and $\frac{\hat{\varepsilon} + b_j}{c((1 - G(\hat{\varepsilon})Q))} = \big(1 - \mu\ (\widehat{D})\big)\sigma p_0$.

When D equals $\widehat{D}$, since $\frac{\hat{\varepsilon} + b_j}{c((1 - G(\hat{\varepsilon})Q))}$ is rising with $\hat{\varepsilon}$ at the stable equilibrium, then higher values of $\widehat{D}$, will cause $1 - \mu$ (D) to rise and $\hat{\varepsilon}$ to rise, and the number of disobedients will decrease as $\widehat{D}$ rises. However, the value of d will rise. An increase in the upper limit on D, a more tolerant situation, will mean a smaller group of disobedients doing a more annoying thing.

Holding $\widehat{D}$ constant, $\mu(\widehat{D})$ is rising with $\varphi$, $K$, and V and falling with$\theta_L$, $\sigma$ and $p_0$, and hence $(1-\mu(\widehat{D}))\sigma p_0$ is falling with $\varphi$, $K$, and V and rising with $\theta_L$, $\sigma$ and $p_0$. Hence the size of the protest is rising with $\varphi$, $K$, and V and falling with $\theta_L$, $\sigma$ and $p_0$. The intensity of the protest, however, since $\widehat{D}$ is fixed will be decreasing in $\varphi$, $K$, and V and rising with $\theta_L$, $\sigma$ and $p_0$.

In this region, the size of the protest will be rising with $b_j$ and Q, but the intensity of the protest will be falling.

If the $\widehat{D}$ constraint doesn't bind, then planners choose d so that $d(1-G(\varepsilon^*))Q - \varphi K + VF\left(\frac{\gamma(\sigma p_0 - \pi_{max})}{1-\pi_{max}} - \theta_L\right) = VF(\gamma - \theta_L)$, where $\varepsilon^*$ and $\pi_{max}$ are fixed by $\pi_{max} = \frac{\varepsilon^* + b_j}{c((1-G(\varepsilon^*)Q)}$ and $\varepsilon^*$satisfies $c((1-G(\varepsilon^*))Q) = -c'((1-G(\varepsilon^*))Q)Q(\varepsilon^* + b_j)g(\varepsilon^*)$. In this case, the number of disobedients is unrelated to $\varphi$, K, V, $\sigma$, $p_0$ and $\theta_L$, since it is fixed by $\pi_{max} = \frac{\varepsilon^* + b_j}{c((1-G(\varepsilon^*)Q)}$. However, d is rising with $\varphi$, K, V, and falling with and $\theta_L$, $\sigma$, and $p_0$. The value of $\pi_{max}$ is rising with both $b_j$ and Q, and hence D must be rising with both of those variables.

*Proof of Proposition 7:* The welfare to the planner from a pooling equilibrium, whether or not the leaders repress or tolerate, equals $B_H F(\gamma p_0 - \theta_L)$. In a fully separating equilibrium the planners welfare equals $B_H(\sigma p_0 F(\gamma - \theta_L) + (1-\sigma p_0)F(-\theta_L))$. IF F(.) is concave, then the $F(\gamma p_0 - \theta_L) > (p_0 F(\gamma - \theta_L) + (1-p_0)F(-\theta_L)$, $F(\gamma - \theta_L) > F(\gamma p_0 - \theta_L)$ and $B_H(\sigma p_0 F(\gamma - \theta_L) + (1-\sigma p_0)F(-\theta_L))$ is monotonically increasing in $\sigma$. Hence there must exist a single value of $\sigma$ between 1 and $1/p_0$ at which $\sigma p_0 F(\gamma - \theta_L) + (1-\sigma p_0)F(-\theta_L) = F(\gamma p_0 - \theta_L)$. Monotonicity implies that the planner will always prefer protest if $\sigma$ is higher than that amount and doing nothing if $\sigma$ is lower than that amount.