Online Appendix
Rumors, Kinship Networks, and Rebel Group Formation

Jennifer M. Larson and Janet I. Lewis

July 12, 2017

Contents

1 Elaborated Model of Civilian Secret-Keeping

1.1 Overview and Relationship to Existing Theories

In the article text and below with greater elaboration, we present a model of civilian behavior in which each civilian chooses whether to provide information about the rebels to the government or not. A network of trusted communication among the civilians determines their access to pertinent rumors about the rebels' capabilities and cause, and helps them form expectations about fellow civilians. Specifically, the network (fixed ex ante) transmits rumors seeded by the rebels which shape civilians' individual valuations of the rebels. Given these valuations, civilians then play a coordination game in which they can keep secrets or inform the government about the rebels. The fragile rebels are particularly susceptible to the first few informants; those who value the rebels highly have a strong incentive not to be the lone informant. If many are expected to inform, though, even those who value the rebels highly have little incentive to turn down the government's payment and keep quiet.

The results generated by the model offer comparative statics at the level of the network and focus on the importance of secrecy in the earliest stages of rebellion. Trusted communication networks allow civilians to spread rumors and coordinate – the more these networks feature low fragmentation and short paths over the area near aspiring rebels, the more likely is full secret-keeping in equilibrium.

Our focus on kinship networks' relevance to rumor transmission, and how those networks map onto ethnic demography, shares the approach of several works that emphasize ethnic demography's importance – especially the geographic concentration of people from a single ethnic group – in facilitating the opportunity for rebellion (Toft, 2002, 2003; Laitin, 2004; Weidmann, 2009). Most emphasize that communication networks can be relevant to recruitment, resource mobilization, and organizational cohesion of mature rebel groups, following the call of Gould (1991) to consider whole network structures. For example, Gubler and Selway (2012) note that the ease of communication, especially within ethnic groups, can facilitate mobilization and that geographic divisions can impede this benefit. Similarly, Petersen (2001) and Cederman, Weidmann and Gleditsch (2011) suggest that preexisting social networks among communities (conceptualized as ethnic groups in the latter work) can help overcome collective action problems that may be present in high-risk participation and recruitment. Staniland (2012) finds that socially divided networks are inferior for enforcing collective action and keeping elite feuding and on-the-ground control problems in check. Parkinson (2013) shows that the presence of informal ties offers access to information, finance and supplies, and that these strong quotidian ties can serve as bridges between formal organizations.

Many of these findings are similar in kind to our results– links in a network offer access to a helpful resource for rebels, and the readier that access, the better the network can facilitate rebellion. Fragmentation of the network, manifested in ethnic, social, class and/or geographic divisions, hinder this process. Our results, however, differ in two important respects.

First, the nature of ties and the function of the network is different in the earliest stages of rebellion considered here. While these earliest stages are a small slice of the lifespan of rebel groups that manage to endure, we focus on them since (as we stress in the article text) they are so rarely examined, yet so important to understanding how organized violence begins. As discussed above, the problem incipient rebels face is not one of encouraging civilians to take public, costly actions that necessarily entail incentives to free ride. The problem is one of remaining clandestine, which requires that nearby civilians are convinced that many other of the nearby civilians are convinced of the promise of the aspiring rebels. Rebels need civilians to perceive widespread support for them.

A key assumption here is that rebel groups are initially small, poor, and lack coercive capacity – and are thus highly fragile and more concerned with remaining in obscurity than they are with swelling their ranks.[1] The latter carries substantial risks; due to the extreme vulnerability of incipient groups, each new member must be carefully screened.[2] Of course, the secrecy-enhancing role of civilian networks in the initial stages of rebellion does not preclude their usefulness in facilitating more public forms of collective action in later stages of rebellion as well.

Second, our results reveal the precise features of networks that bear on rebel viability. Fragmentation is indicated by social divisions, but can be measured more precisely using Definition 1 below. Even an unfragmented network can be better or worse for rebel viability depending on whether it is a low- or high- distance network. Further, according to our results and contrary to common assertions, networks need not be "dense" to facilitate rebel group formation. This model joins the games-on-networks literature which incorporates precise network structures into models of strategic interaction (Gould, 1993; Jackson and Wolinsky, 1996; Chwe, 2000; Dixit, 2004; Bramoullé and Kranton, 2007; Siegel, 2009; Lippert and Spagnolo, 2011; Patty and Penn, 2014; Wolitzky, 2013; Larson, 2016), often by specifying who plays whom (see, e.g., Bramoullé and Kranton, 2007) or by specifying who learns about whom (see, e.g., Larson, 2017). Here the network transmits relevant information *before* the game that will be used during the game, and can serve as a coordination device within the game. The theoretically-grounded network measures found to matter pertain to spreading information and coordinating action, and may prove useful for future research into the later stages of rebellion as well.

1.2 Model Setup

Here we reintroduce the model that is presented in the article text, elaborating on assumptions underlying the strategic environment and the game setup.

[1] As explained in the article text, while material resources bear on rebel trajectories, it is rare that rebel groups are "born" with access to substantial financing via natural resources. A similar logic applies to the external sponsorship of rebel groups. While states often delegate war to rebel groups, with important consequences for groups' behavior (Salehyan, 2010; Salehyan, Siroky and Wood, 2014), especially since the end of the Cold War it is unusual to find external sponsors building proxy armies from the ground up; they are not likely to finance a group until that group has demonstrated basic coercive capacity.

[2] Groups may move through this phase more quickly if they are fortunate enough to have certain special endowments, such as an easily-accessed, relatively large core of highly-trusted fighters from an earlier episode of organized violence (Zukerman-Daly (2012); Finkel (2015)).

Consider a set N of n individuals (here, "civilians"). Civilians have a fixed and exogenously-given "trusted communication network" defined by the pair (g, N) with $n \times n$ adjacency matrix g where $g_{i,j} = g_{j,i} = 1$ indicates a link between $i \neq j \in N$. To simplify notation, we will refer to the network as g. Links in the networks are undirected and unweighted, and the network is common knowledge to individuals in N.[3]

Before the game begins, a rebel group begins operating and all civilians within the area learn some information about the group.[4] While rebels of course do not share important tactical information with those outside of their inner circle, locals will inevitably learn basic information about their existence, identities, and general whereabouts, which is sufficient to damage the rebel group if the government knew it.[5] Let the probability that the rebels are ultimately successful, $p(\#R)$, be decreasing in the number of informants $(\#R)$ such that the marginal impact of each additional informant is decreasing.

Suppose the rebels seek out a trusted contact $i^{seed} \in N$ from among the civilians and provide him with a framed, compelling message about the rebels, which may include goals, promises, glowing assessments of future capabilities, arguments for why rebellion is just, etc. Note that this framed account is about aspects of the rebels that are generally prospective or impossible to confirm; thus we call it a *rumor*, whereas aspects of the rebellion about which the government seeks to learn are factual; we call this *information*. We assume all civilians in the area where rebels form learn information that could damage the rebels if shared with the government, but only those connected via the trusted network g receive and are potentially influenced by the rumor. Let B be the benefits that i^{seed} expects the village would receive if the rebels are successful after hearing the rumor.[6] The rumor spreads through the network as a message containing i^{seed} and B, losing potency so that the benefit that any message recipient i expects to gain from the rebels if they succeed, b_i, is a function of the network distance the message spreads before reaching i and of the seed's valuation: $B\epsilon^{\ell(seed,i)}$.[7] If the message does not reach a player j (because $\ell(seed, j) = \infty$), then $b_j = 0$. The degradation process and ϵ are common knowledge to civilians in N.

[3]The model is agnostic to how or why the trusted links between civilians formed. In the article text, we note that in Subsaharan Africa, shared kinship is an important source of trusted links.

[4]The relevant set of civilians, and hence the relevant network, is defined by geographic proximity to the aspiring rebels. This is in contrast to other models in which the network is defined by membership in some group– student in a class, member of a committee– and has the interesting consequence that networks could be favorable to rebels in some very small space, but the rebels operate in a larger space. When zoomed out to that space in which the nascent rebels operate, networks may be unfavorable.

[5]Civilians may do so, for example, by observing training exercises, identifying some rebels, detecting the location of bases, etc. Merely confirming the existence of a nascent rebel group can be valuable to the government which, early on, may be trying to discern whether a nascent rebel organization truly exists.

[6]Note also that this draws on the fact that rebels often have a personal contact in the village that they use to insert the rumor into a village. We argue that such a civilian is likely to be persuaded by the rebel's message, though we allow *how* persuaded to vary.

[7]$\ell(seed, i)$ is the number of links in the shortest path between i^{seed} and i. Degradation is a standard feature of information flowing through a network. The model is agnostic about the reason: the message may become more error-prone as it passes from person to person as in the game of telephone, it may be less believable as it extends farther from the source, or it may resonate less with people as it travels a greater social distance.

After the initial spread of rumors seeded by the rebels, the game begins. In the game, the government arrives in the area[8] and asks all civilians individually and simultaneously about the rebels. A civilian i chooses an action $r_i \in \{0, 1\}$: inform the government about the rebels ($r_i = 1$) or keep the information secret ($r_i = 0$). The government offers each civilian $\gamma > 0$ for information (choosing $r_i = 1$), 0 for silence (choosing $r_i = 0$). Call M_x the marginal impact of an additional informant on the probability of rebel success given x informants so that $M_x = p(x) - p(x + 1)$. By the assumptions on p above, an individual has the greatest marginal impact on the rebels' success when he is the sole informant. This captures the context in which rebel success is sensitive to a small number of informants.[9] Civilian i taking action r_i earns payoff

$$u_i(r_i) = b_i p(R_{-i} + r_i) + r_i \gamma,$$

where $R_{-i} = \sum_{j \neq i} r_j$.

1.3 Secret-Keeping in Equilibrium

In this one-shot simultaneous game, an equilibrium is a subdivision of N into a (possibly empty) subset S of those who choose $r_i = 0$ (keeping secrets) and a (possibly empty) subset R of those who choose $r_i = 1$ (inform) such that $S \cup R = N$, $S \cap R = \varnothing$, and no one in either subset has an incentive to switch their action given the partition.

An equilibrium with at least $n^\star$ secret-keepers exists only if $\exists\ b^\star$ such that

$$b^\star M_{n-n^\star} = \gamma, \text{ and}$$
$$\#\{i \in N | b_i \geqslant b^\star\} \geqslant n^\star.$$

The benefit of keeping secrets depends on the number of others who will keep secrets as well. To have an equilibrium with $n^\star$ secret-keepers, there must be at least $n^\star$ civilians who prefer to keep secrets given that $n^\star - 1$ other civilians will keep secrets. With this logic, we can specify the necessary and sufficient conditions for an equilibrium with exactly $n^\star$ secret-keepers.

Proposition 1 (Equilibrium Conditions). *An equilibrium with exactly $n^\star$ secret-keepers*

[8] In the article text, we explain that a typical feature of weak states is the lack of state presence or any intelligence apparatus in villages away from the capital city. To extract information about a prospective rebel group, the government must send emissaries to the area.

[9] That M is positive and potentially large means that civilians may not have a dominant strategy of informing. Rather than exhibiting classic free riding incentives, this setup can produce herding– civilians have the greatest incentive to keep secrets when everyone else is expected to keep secrets as well, all else equal. Our aim is to capture the fragility of early rebel groups.

exists iff $\exists$ *a pair* $(b_s^\star, b_r^\star)$ *such that*

$$\begin{aligned} b_s^\star M_{n-n^\star} &= \gamma, \\ b_r^\star M_{n-n^\star-1} &= \gamma, \\ \#\{i \in N | b_i \geqslant b_s^\star\} &= n^\star, \text{ and} \\ \#\{i \in N | b_i \leqslant b_r^\star\} &= n - n^\star. \end{aligned}$$

Proof. Consider a partition of N into S, the set of civilians playing $r_i = 0$, and R, the set of civilians playing $r_i = 1$. Call x the size of S: $x = \#S$. This partition is an equilibrium iff no civilian has an incentive to switch his action given the actions of others. Civilians $i \in S$ earn $b_i p(n-x)$ from playing $r_i = 0$; deviating would earn $b_i p(n-x+1) + \gamma$. So long as $b_i(p(n-x) - p(n-x+1)) = b_i M_{n-x} \geqslant \gamma$, no $i \in S$ has an incentive to deviate. Civilians $i \in R$ earn $b_i p(n-x) + \gamma$ from playing $r_i = 1$; deviating would earn $b_i p(n-x-1)$. So long as $b_i(p(n-x-1) - p(n-x)) = b_i M_{n-x-1} \leqslant \gamma$, no $i \in R$ has an incentive to deviate. Hence, if for some value of x there exists a pair $(b_s^\star, b_r^\star)$ defined as in Proposition 1 for a partition $\{S, R\}$, the partition is an equilibrium with x secret-keepers. Now suppose there is a partition $\{S, R\}$ that is an equilibrium with $n^\star$ secret-keepers but there is no such $(b_s^\star, b_r^\star)$. Then either there is no such $b_s^\star$ or no such $b_r^\star$. Without loss of generality, suppose there is no such $b_s^\star$. Then it must be that there exists at least one civilian playing $r_i = 0$ in equilibrium for whom $b_i < b_s^\star$. But then $b_i M_{n-n^\star} < \gamma$, a contradiction. The same logic holds for $b_r^\star$. Hence if $\{S, R\}$ is an equilibrium with $n^\star$ secret-keepers, there must be a pair $(b_s^\star, b_r^\star)$ with the properties in Proposition 1 which characterize it. □

The first and third relationship of Proposition 1 ensure that no civilian currently keeping secrets would do better by informing given everyone else's actions, and the second and fourth ensure that no civilian currently informing would do better by informing given everyone else's actions.

It follows that an equilibrium in which *every civilian* keeps secrets exists iff $\exists$ $b^\star$ such that

$$b^\star M_0 = \gamma, \text{ and}$$
$$\#\{i \in N | b_i \geqslant b^\star\} = n.$$

The conditions can be combined into the following succinct expression:

Corollary 1 (Full Secret-Keeping). *An equilibrium in which every civilian keeps secrets exists iff*

$$b_i M_0 \geqslant \gamma$$

$\forall i \in N$.

The existence of a full secret-keeping equilibrium depends on the set of civilian valuations of the rebels $\{b_i\}_{i \in N}$, the size of the government payment for information γ, and the marginal impact of a single informant M_0 on the rebels' ultimate probability of success. The greater the lowest valuation of the rebels among the civilians, the smaller the government's offer, and the more susceptible the rebels are to even a single informant, the more likely the full secret-keeping equilibrium is to exist.

The full secret-keeping equilibrium need not be unique. Next we turn to the issue of multiple equilibria.

1.4 Equilibrium Selection

When all civilians' valuations of the rebels are sufficiently high, the only equilibrium is the full secret-keeping equilibrium. Specifically, the full secret-keeping equilibrium is unique if and only if

$$b_i M_{n-1} > \gamma \tag{1}$$

for all $i \in N$. In this trivial case, each civilian values the rebels so highly that she regards the net impact she has on the rebels' ultimate success to be greater than the government's payment *even if all $n-1$ other civilians inform the government.* In this case, her valuation compels her to keep secrets independent of what other civilians plan to do.

However, a civilian need not value the rebels so highly for full secret keeping to be an equilibrium. Because secret-keeping is more valuable when more other civilians plan to keep secrets too, at lower values of b, a person's expectation that others will keep secrets can make secret-keeping her best action.

For smaller values of b, full secret keeping can still be an equilibrium, but other equilibria can exist too. In this case, so long as $b_i M_0 \geqslant \gamma$ (as per Corollary 1), full secret-keeping will be an equilibrium, though need not be unique. Since, by assumption, $M_0 > M_{n-1}$, smaller values of b can support full secret keeping in equilibrium because the expectation that others will keep secrets increases the utility of a civilian's choice to keep secrets.

The game admits multiple equilibria when there are individuals who value the rebels highly enough to keep secrets given that x others do as well, but not so highly that if fewer others keep secrets, they gain from informing too. This is possible when p is rapidly decreasing in $\#R$. In that case, the marginal impact from refraining from informing becomes very small and the probability of attaining any benefits from successful rebels becomes so low, equilibria with little- or no-secret keeping are possible as well. In such a case, the question becomes which equilibrium is played by civilians.

As a first cut, we restrict attention to the equilibria with the largest number of secret keepers possible, the "maximal secrecy equilibria." These equilibria are of interest not only because they correspond to rebel groups having the greatest chance at attaining viability, but also because they have two desirable properties: they are Pareto efficient, and they are universally individually preferred by all civilians to all other equilibria.

We show that for the setup presented here, when there are multiple equilibria, not only is the group better off in aggregate when the group selects the equilibrium with the greatest number of secret-keepers, but each individual is also personally better off when the group selects this equilibrium over any other equilibrium that exists.

Lemma 1. *When multiple equilibria exist, the equilibrium with the largest number of secret-keepers yields the greatest total welfare.*

The proof of Lemma 1 is combined with the proof of the following related Lemma:

Lemma 2. *When multiple equilibria exist, a civilian earns the greatest payoff in the equilibrium with the largest number of secret-keepers compared to any other equilibrium.*

Proof. An equilibrium is characterized by a set of secret-keepers S and a set of informants R, where $S \cup R = N$. For a set of civilians and a set of model parameters, order the set of all equilibria by number of secret-keepers. Label the set of secret-keepers in the equilibrium with the most secret-keepers S, and call this equilibrium EQ. Without loss of generality, consider any equilibrium EQ' with strictly fewer secret-keepers and label the set of secret-keepers in this equilibrium S'. By the equilibrium conditions in Proposition 1, for any equilibrium with $n^\star$ secret-keepers and $n - n^\star$ informants among a set of civilians with individual benefits $\{b_i\}_{i \in N}$, $b_i \leqslant b_j \ \forall i \in R, j \in S$. Hence, there are three sets of civilians to compare across the two equilibria: (1) Those in $S \cap S'$, (2) those in $R \cap R'$, and (3) those in S in EQ who fall in R' in EQ'. In set (1), secret-keepers $j \in S \cup S'$ earn $u_j(0) = b_j p(\#R)$ under EQ and $u_j(0) = b_j p(\#R')$ under EQ'. Since $\#R < \#R', p(\#R) \geqslant p(\#R')$, and so $u_j(0|EQ) \geqslant u_j(0|EQ') \forall j \in S \cap S'$. In set (2), informants $i \in R \cup R'$ earn $u_i(1) = b_i p(\#R) + \gamma$ under EQ and $u_i(1) = b_i p(\#R') + \gamma$ under EQ', so again $u_i(1|EQ) \geqslant u_i(1|EQ') \ \forall i \in R \cap R'$. Those k in set (3) earn $u_k(0) = b_k p(\#R)$ under EQ and $u_k(1) = b_k p(\#R') + \gamma$ under EQ'. By the equilibrium conditions, it must be that $\gamma \leqslant b_i M_{\#R} \ \forall i \in S$, which implies that, $\gamma \leqslant b_i(p(\#R) - p(\#R'))$, which implies that $b_i p(\#R') + \gamma \leqslant b_i p(\#R)$. Hence, $\forall \ k \in S$ in EQ who fall in R' in EQ', $u_k(0|EQ) \geqslant u_k(1|EQ')$. Hence, all individuals have weakly greater utility under EQ than EQ' (Lemma 2). It follows that the sum of individual utilities is weakly greater under EQ than EQ' (Lemma 1). □

In short, when there exist multiple equilibria, the maximal secrecy equilibrium is the strictly-preferred Pareto efficient Nash equilibrium. When a full secret-keeping equilibrium exists, it is the strictly-preferred Pareto efficient Nash equilibrium.[10]

1.5 Note on Risk Dominance

A concern with Pareto efficiency as an equilibrium selection criterion in a one-shot game is that other equilibria may be risk dominant, which may be a more compelling reason for players to play them (Harsanyi, 1995). For instance, in the classic stag hunt game, if players have more to lose by shooting at the stag when others shoot at the hare compared to their expected loss if they shoot at the hare instead of the stag, players may reasonably prefer to

[10]This is different from merely saying that certain equilibria are efficient. Here, maximal secret-keeping is not only globally efficient, but each individual prefers this equilibrium (if it exists) as well.

shoot at the hare despite the Pareto efficiency of the equilibrium in which all shoot at the stag. Here, we may worry that keeping secrets when others inform is worse than informing when others keep secrets, in which case civilians may prefer to inform despite the Pareto efficiency of the equilibrium in which all keep secrets.

To explore risk dominance in this case, consider a simple version of the game above in which there are only two civilians, i and j. Their payoffs given all combinations of actions are as follows:

		Player i	
		Secret	Inform
Player j	Secret	$b_i p(0)$, $b_j p(0)$	$b_i p(1)$, $b_j p(1) + \gamma$
	Inform	$b_i p(1) + \gamma$, $b_j p(1)$	$b_i p(2) + \gamma$, $b_j p(2) + \gamma$

Assume that

$$b_i M_0 \geqslant \gamma \text{ and } b_j M_0 \geqslant \gamma \tag{2}$$

so that (Secret, Secret) is a Nash equilibrium. When $b_i M_1 \geqslant \gamma$ and $b_j M_1 \geqslant \gamma$ also hold, (Secret, Secret) is the only Nash equilibrium. Suppose instead that

$$b_i M_1 < \gamma \text{ and } b_j M_1 < \gamma. \tag{3}$$

Now (Inform, Inform) is also a pure strategy Nash equilibrium.

Since $p(2) < p(0)$ and M is decreasing in the number of informants, (Secret, Secret) is the strictly–preferred Pareto efficient Nash equilibrium (since $b_i p(0) > b_i p(1) + \gamma$, $b_i p(0) > b_i p(2) + \gamma$ and likewise for j).

Now we can ask: when is (Secret, Secret) also risk dominant? Comparing the product of gains from deviating from the (Secret, Secret) equilibrium to the product of gains from deviating from the (Inform, Inform) equilibrium, (Secret, Secret) risk dominates (Inform, Inform) when:

$$\begin{aligned}
(b_i p(1) - (b_i p(2) + \gamma))(b_j p(1) - (b_j p(2) + \gamma)) &< (b_i p(1) + \gamma - b_i p(0))(b_j p(1) + \gamma - b_j p(0)) \\
\Leftrightarrow (b_i M_1 - \gamma)(b_j M_1 - \gamma) &< (b_i(-M_0) + \gamma)(b_j(-M_0) + \gamma) \\
\Leftrightarrow (b_i M_1 - \gamma)(b_j M_1 - \gamma) &< (b_i M_0 - \gamma)(b_j M_0 - \gamma)
\end{aligned}$$

By Assumption 2 above, $(b_i M_0 - \gamma) \geqslant 0$ and $(b_j M_0 - \gamma) \geqslant 0$. Likewise, by Assumption 3, $(b_i M_1 - \gamma) < 0$ and $(b_j M_1 - \gamma) < 0$. Consequently, $(b_i M_0 - \gamma) > (b_i M_1 - \gamma)$ and

$(b_j M_0 - \gamma) > (b_j M_1 - \gamma)$. It follows that the above inequality holds only if either

$$\begin{array}{rcl} |b_i M_1 - \gamma| & < & (b_i M_0 - \gamma) \\ & \text{or} & \\ |b_j M_1 - \gamma| & < & (b_j M_0 - \gamma) \end{array}$$

which is to say that the condition does not hold if both inequalities are violated. By Assumption 3, the inequalities are equivalent to

$$\begin{array}{rcl} \gamma - b_i M_1 & < & (b_i M_0 - \gamma) \\ & \text{or} & \\ \gamma - b_j M_1 & < & (b_j M_0 - \gamma). \end{array}$$

Rearranging, (Secret, Secret) risk dominates (Inform, Inform) only if

$$\frac{2\gamma}{M_0 + M_1} < b_i \text{ or } \frac{2\gamma}{M_0 + M_1} < b_j.$$

Since by definition $M_x = p(x) - p(x+1)$, we can rewrite this condition as

$$\frac{2\gamma}{p(0) - p(2)} < b_i \text{ or } \frac{2\gamma}{p(0) - p(2)} < b_j. \tag{4}$$

Recall that (Secret, Secret) and (Inform, Inform) are both pure strategy Nash equilibria in the 2 player game when both players value the rebels in the following range:

$$\frac{\gamma}{M_0} < b_i < \frac{\gamma}{M_1}. \tag{5}$$

Call the highest valuation of either player $b^{max} = \max\{b_i, b_j\}$. The conditions above value implies a cut of the interval in (5) such that valuations at the low end result in (Inform, Inform) being risk dominant, and valuations at the high end result in (Secret, Secret) being risk dominant. Specifically, (Inform, Inform) is the risk dominant equilibrium when

$$\frac{\gamma}{M_0} < b^{max} < \frac{2\gamma}{M_0 + M_1} \tag{6}$$

and (Secret, Secret) is the risk dominant equilibrium when

$$\frac{2\gamma}{M_0 + M_1} < b^{max} < \frac{\gamma}{M_1}. \tag{7}$$

In other words, in the two person version of the game presented here, when (Secret, Secret) and (Inform, Inform) are both equilibria, (Secret, Secret) is Pareto efficient and can also be strictly risk dominant when Condition 7 is satisfied. When the government is able to make a very large offer in exchange for information, and when the difference in rebel success when everyone informs versus when no one informs is small, this condition is harder to satisfy.

This condition for the two person game, and the extension of its logic to a multiple person version of the game, has important takeaways for the scope conditions. Insofar as risk dominance is a good predictor of behavior– if indeed people only play strategies that insure them against the risk that the other player will deviate– then the results presented here are unlikely to inform two specific contexts:

1. High capacity states. States that can amass large amounts of resources and offer lucrative bonuses (and can ensure that they will be delivered) in exchange for information are likely to offer large γ. In these cases, the value of this offer may overwhelm the impact on expected rebel success, making (Inform, Inform) the risk dominant equilibrium even though (Secret, Secret) is the Pareto efficient equilibrium.[11] To see this, note that the region in which (Secret, Secret) is risk dominant– shown in (7)– shrinks relative to the region shown in (6) in which (Inform, Inform) is risk dominant.

2. Rebel groups that are insensitive to multiple information revelations. When rebel groups are insensitive to informants, so that the difference between many and few informants bears little on their probability of success, (Inform, Inform) can be risk dominant. Based on the evidence presented in the body of the paper, we argue this is unlikely to be the case for nascent rebel groups. Groups that are in more advanced stages of formation, on the other hand, or that have an initial advantage that makes them less vulnerable, like secure access to resources or partnerships with more advanced groups (which we point out was not the case for incipient rebels in Uganda), may be less sensitive to the difference between none and some informants. In such a case, (Inform, Inform) may be the risk dominant equilibrium.[12]

On the other hand, so long as the benefit that civilians expect to receive is large relative to the government's offer and the rebels are sufficiently sensitive to informants, keeping secrets is both Pareto efficient and risk dominant.

[11] And of course, as the conditions above make, particularly high capacity states may offer such a large amount that secret-keeping is not an equilibrium outcome.

[12] Note that (Inform, Inform) is an equilibrium for more values of b when M_1 is small; that is, when the marginal impact of a second informant is low.

1.6 Note on Coordination on Equilibrium

Preferring an equilibrium is different from having the means of coordinating on it. In the context considered here, though, individuals have a tool useful for coordination at their disposal: their network.

Common knowledge of the seed, B, and the network is sufficient for everyone to know the private valuations of everyone else in a component of the network. We might also imagine that civilians with high enough b_i who have an interest in secret-keeping turn to their trusted contacts to actively discuss options and try to generate coordination. Whether the process is inferential, or an active use of trusted ties to form coordination, the network determines how successful this coordination can be.[13]

1.7 Characterizing the Role of the Network

Our setup assumes that the model transmits rumors that the rebels seed about themselves, which affects the civilians' valuations of them. In order to characterize network features that matter for secret-keeping, we will focus on the consequences of different network structures for the civilians' valuations of the rebels (b_i).

The analysis thus far implies a correspondence between the distribution of values of b and full secret keeping equilibria, summarized in Table 1.

Highest values of b_i ($\min_{i \in N}\{b_i\} > \frac{\gamma}{M_{n-1}}$)	Full secret-keeping equilibrium unique, civilians don't care what other civilians do
Next highest values	Full secret-keeping equilibrium exists, depends on other civilians keeping secrets too, Pareto Efficient and Risk-Dominant
Next highest values	Full secret-keeping equilibrium exists, depends on other civilians keeping secrets too, Pareto Efficient but NOT Risk-Dominant
Low values ($b_i M_0 < \gamma$ for some i)	Full secret-keeping NOT equilibrium

Table 1: The consequences of the set of b_i for $i \in N$ for full-secret keeping in equilibrium.

[13]The difficulty of coordinating with civilians outside of a network component is another reason why fragmented networks can undermine secret-keeping (and rebel success), discussed below. If civilians know that there is only one message, the network is fragmented, and all start with $b_i = 0$, then civilians know that those outside a component will always choose $r_i = 1$, making the choice of $r_i = 0$ less valuable even among those in the same component. If civilians don't know the priors of those outside their component and/or don't know that rebels choose only one seed, they are left to guess. Their level of optimism about others with whom they never communicate determines the extent of barriers to coordination.

To explore the consequences of different network structures, we will focus on their effect on the distribution of b_i $\forall i \in N$, which affects. Since $b_i = B\epsilon^{\ell}(seed, i)$, some networks are more conducive to full secret-keeping than others.

1.8 Network Comparative Statics

We report the consequences of different network structures for the distribution of b_i, the civilians' valuations of the rebels, because this has consequences for full secret keeping in equilibrium. We focus on the extent to which full secret keeping is more likely in equilibrium, in the sense that the condition for secret keeping by all civilians becomes easier to satisfy.[14]

Secret-keeping is more valuable for a civilian when she values the rebels highly (concerns b), and knows that many others do too (concers M). Since the potency of pro-rebel messages decays as they spread further through the network ($b_i = B\epsilon^{\ell}(seed, i)$), networks that transmit the rebels' rumors to all civilians through few steps are optimal for incentivizing full secret-keeping.

One feature of networks that directly bears on this transmission is the extent to which a network is separated into different components. This matters because civilians residing in a different component than the one into which the rebels inserted their rumor never hear it. For these civilians, $b_i = 0$ and $\gamma > 0$, and civilians in the informed components know that these civilians are uninformed. We then have a straightforward relationship between networks with multiple components and the maximum number of secret-keepers in equilibrium.

Call a network's "fragmentation" the proportion of nodes not in the largest component:

Definition 1 (Network Fragmentation). *Enumerate a network's k components $C_1 \dots C_k$ in descending order of size so that C_i is the set of all nodes in the ith component, and C_1 is the set of nodes in the largest component. A network g's fragmentation can be written*

$$Frag(g) = \sum_{i=2}^{k} \frac{\#C_i}{n}$$

for $k > 1$ and Frag(g) = 0 for $k = 1$.

In other words, connected networks[15] have no fragmentation. Networks with multiple components are most fragmented when the components are the same size. Now the following comparative static follows immediately: networks that are fragmented result in informants in equilibrium; the more fragmented a network, the smaller the maximum number of secret-keepers.

[14] We could also specify equilibria with partial secret-keeping, and report consequences for a "maximal secrecy equilibrium." The results presented below that make the conditions for full secret keeping easier to satisfy also make the conditions for a partial secret-keeping equilibrium easier to satisfy for the secret-keepers and maximize the size of the largest possible set of secret-keepers in equilibrium.

[15] A connected network is one in which there exists a finite path between any pair of nodes– not to be confused with a complete network in which every possible link is present.

Proposition 2 (**Fragmentation Undermines Full Secret-Keeping).** *For a network g among a group of civilians N, if $frag(g) > 0$, a full secret-keeping equilibrium does not exist.*

The proof of this proposition is straightforward: since all civilians outside of the component that the rebels contact have $b_i = 0$, and since $\gamma > 0$, all these civilians strictly prefer informing the government. So long as there is one such civilian (as there is for any positive value of network fragmentation), full secret keeping is not an equilibrium outcome. Moreover, fragmentation also affects the maximum number of secret-keepers in not-full-secret-keeping equilibrium. Consider the best case scenario for rebels: their trusted contact lies in the largest component of the network. Even in this case, all outside the component are outside the reach of their message. Any equilibrium must have at least all civilians outside the component with the rumor choosing inform; other additional civilians may inform too.

Corollary 2. *The larger the value of Frag(g), the larger the minimum number of informants in the maximal secrecy equilibrium (and hence the lower the probability of rebel success p).*

The more fragmented the network, the more civilians have $b_i = 0$ since they cannot be reached by the rebels' rumors. These civilians will not keep secrets in equilibrium since $\gamma > 0$. The more fragmented the network, the more civilians are sure to be informants. Hence, fragmentation places a lower bound on the number of informants in equilibrium. Depending on the valuation of civilians inside the component that the rebels contact and the network structure interconnecting them, they may inform or not. In order to be incentivized to keep secrets they must find the rebels sufficiently valuable to refrain from informing even given that not all of their fellow civilians will also keep quiet. In other words, fragmentation potentially reduces the number of secret-keepers through a second mechanism, by reducing the number of other secret keepers that civilians reached by rebels' rumors expect.

Two networks with the same level of fragmentation share a minimum number of informants in equilibrium– neither can have fewer informants than this floor. However, they may differ in their total maximum number of informants based on the network structure within the component that the rebels contact. This comparison depends on network distance– how many links the rumors seeded by rebels must hop to reach each civilian.

Because the full secret-keeping equilibrium is more likely when b_i is large for all i, the fewer links that news must travel from the rebels' contact among the civilians (the seed) the better for secret-keeping. The following results are stated for connected networks. For fragmented networks, the results pertain to the component in which the seed resides.

Although this set of results pertains to network distances, the precise statistic which encodes network distance depends on the way rebels are able to select their seed from among the civilians.

Some seeds would be better for the rebels than others. If rebels had full choice of their seed and knew the network, they could strategically select the best possible seed from among the civilians. This is the case presented in the article text. On the other hand, rebels may be constrained in their choice; perhaps they only have personal ties to few villagers and these villagers may be a suboptimal choice as seed. We can establish the conditions for the best and worst possible seed in order to bound the consequences.

Proposition 3 (**Diameter and the Worst Possible Seed).** *In a connected network (component), given γ, B, and ϵ, the worst possible seed (in terms of satisfying the condition for full secret-keeping) is one that requires the longest shortest path in the network (component) to reach all other civilians.*

Proof. In a given connected network, given γ, B, and ϵ, the condition for full secret-keeping binds for the civilian i^{min} with the smallest valuation of the rebels: $\min_i\{b_i\}$. The smaller is $b_{i^{min}}$, the more difficult the condition is to satisfy. $b_i = B\epsilon^{\ell(i,seed)}$ takes its minimum when $\ell(i, seed)$ takes its maximum. Hence, the seed that minimizes $b_{i^{min}}$ is the seed that maximizes the distance that a rumor must travel: $\arg\max_{seed}\{\max_i\{\ell(i, seed)\}\}$. □

This proposition implies that, all else equal, the smaller the diameter of the network (component), the better the worst possible seed is for complete secret-keeping.

Corollary 3. *Given γ, B, and ϵ, if the longest shortest path in network g is shorter than the longest shortest path in network g', then if the rebels choose their worst possible seed, the condition for full secret-keeping is easier to satisfy in g than in g'.*

The rebels' trusted contact may be a contact for reasons other than spreading news of a rebellion– the rebel and the seed may be relatives, they may have conducted business together in the past, etc.– and so the rebels may be unlucky in where their contact in the community is situated within the network. The network diameter bounds how unlucky they could be. On the other hand, perhaps rebels have strategically forged a relationship with someone in the community based on his network position or are simply lucky in where their one contact is situated within the network. A slightly different feature of the network determines just how lucky they could be.[16]

Define a network's "time to saturation," a measure related to diameter but instead considers the shortest longest path required to reach all other nodes from a single node.

Definition 2 (**Time to Saturation).** *The Time to Saturation of a network g is the length of the shortest path that a rumor originating with one node would need to traverse to reach any other node:*

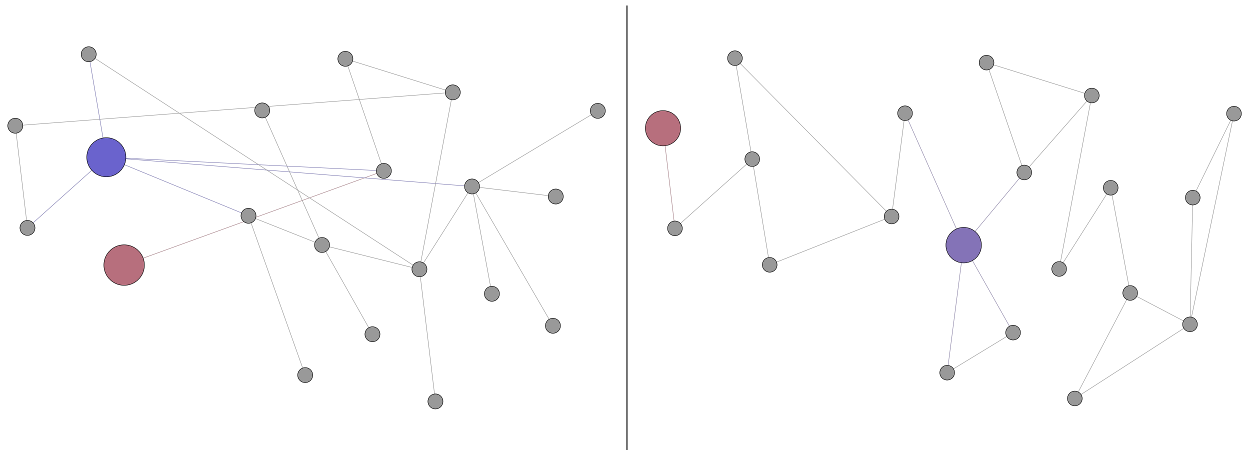
$$TTS(g) = \min_i\{\max_j\{\ell(i, j)\}\}$$

where $\ell(i, j)$ is the length of the shortest path between nodes i and j.

Now we can state the next result:

Proposition 4 (**Time to Saturation and the Best Possible Seed).** *In a connected network (component), given γ and B, the best possible seed (in terms of satisfying the condition for full secret-keeping) is one with the shortest path to all other nodes.*

[16]Relatedly, if the rebels are from the area, they may have a few trusted ties to choose from. The best and worst possible seeds bound how helpful their seed could possibly be.



(a) Low-Distance Network (b) High-Distance Network

Figure 1: Best possible seed shown in blue, worst possible seed shown in red.

The proof relies on the same logic as the proof of Proposition 3. Because the condition for full secret-keeping binds for the player with the minimum valuation, it binds for the player farthest from the seed in the network. The best possible seed is the one that minimizes this maximum distance to every civilian. This result also implies a network feature that matters:

Corollary 4. *The lower the minimum time to saturation, the better the best possible seed is for complete secret-keeping.*

Putting these results together, Figure 1 shows the best possible seed and the worst possible seeds in two example networks.

1.9 Small Network Differences Compound into Large Consequences for Rebels

It is worth noting that small differences in network distances can result in large differences in the number of secret-keepers in equilibrium because network distance compounds the error rate quickly. Figures 2 and 3 show the distribution of support (the distribution of b_i) when the rebels seek out their best possible seed in each network with the original persuasiveness set at $B = 1$ for simplicity, and the figure after shows the same when the rebels use their worst possible seed. In each figure, four values of the error rate ϵ are displayed. The lower the level of support, the easier it is for the government to buy information and the less likely the civilians are to coordinate on keeping secrets. In short, the more mass to the left of 1, the less likely is full secret-keeping in equilibrium.

The above results help to pin down comparative statics at the level of the network. It is important to note that both features should be considered in tandem. While one network that is more fragmented than another has a larger *minimum* number of informants in equilibrium, whether it will in fact have a larger number of informants depends on the network distance

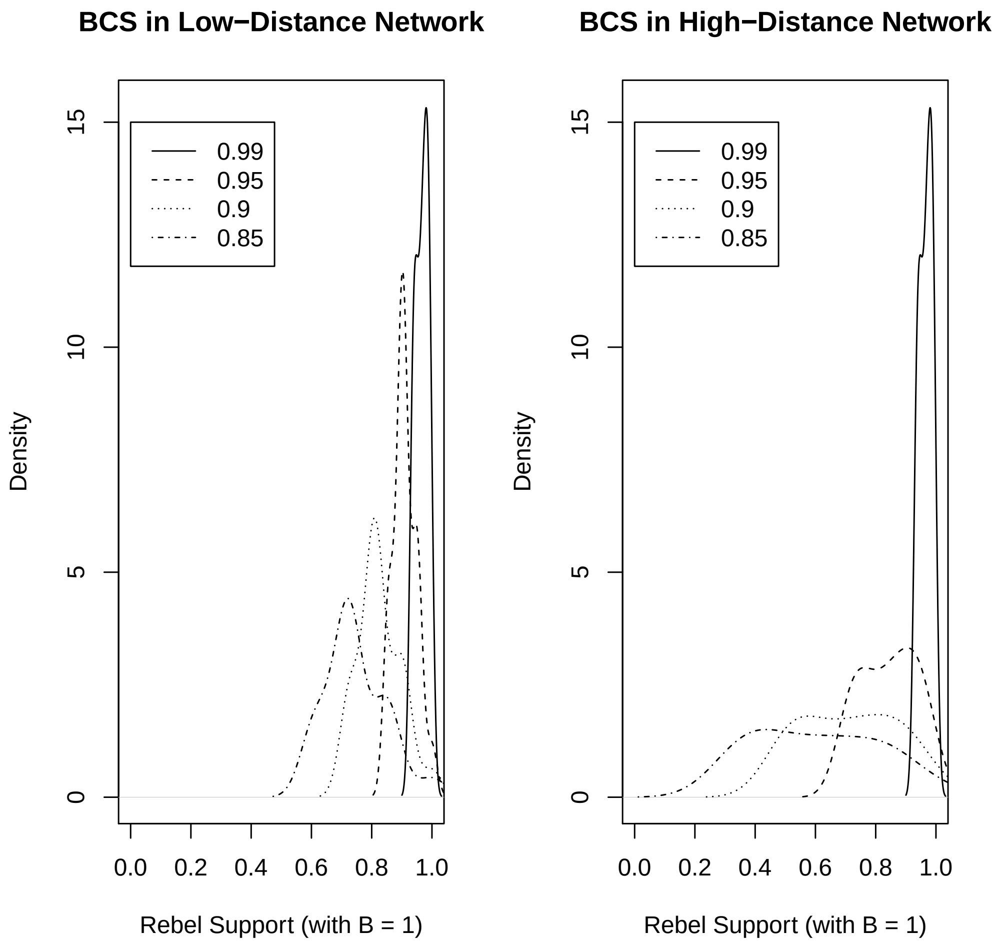


Figure 2: Distribution of support when rebels use the best possible seed, which is colored blue in the high- and low-distance networks in Figure 1. High distances quickly attenuate private valuations, even when the degradation rate is low.

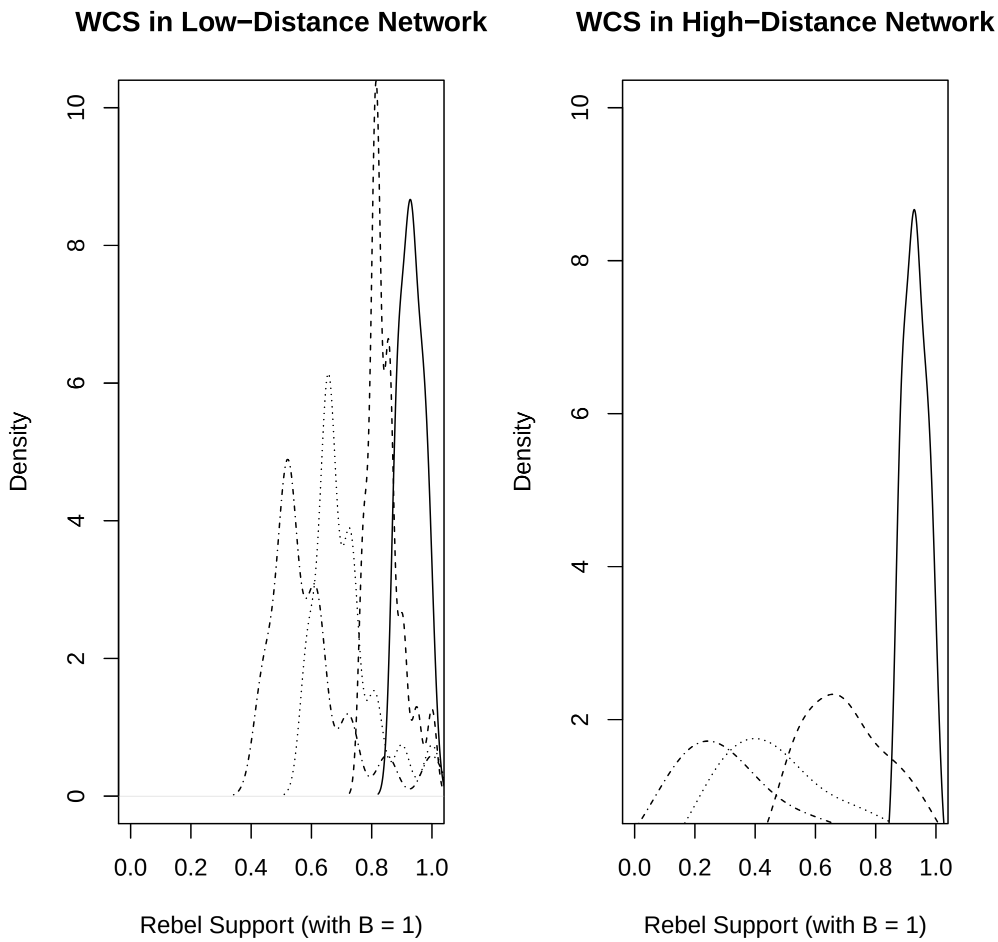


Figure 3: Distribution of support when rebels use the *worst* possible seed, which is colored red in the high- and low-distance networks in Figure 1. High distances become an even greater problem.

within the component that the rebels contact. Table 2 helps to clarify expectations at the level of the network:

		Network Distance: High	Network Distance: Low
Fragmentation	High	Worst case for rebels	At least some informants
	Low	Possibly some informants	Best case for rebels

Table 2: When fragmentation is high (in fact non-zero), full secret-keeping cannot occur in equilibrium. Network distance within the component that rebels contact determines just how many informants there will be in the equilibrium with the most secret-keepers. When fragmentation is zero, full secret-keeping is possible, and network distance determines how many can be incentivized to keep secrets (and hence how likely full secret-keeping is in equilibrium). Networks that are highly fragmented with high distances within components are the worst scenario for rebels; networks that are unfragmented with low distances are most favorable to full secret-keeping.

1.10 Summary of Comparative Statics

Table 3 summarizes the comparative statics with respect to the ultimate probability of rebel success.

1.11 Note on Beliefs

In the model, the transmission of rumors is neither strategic nor Bayesian. This modeling choice follows standard representations of information spread through a network. Information aggregation or learning on a network presents formidable barriers to pure Bayesian updating (see, e.g., Jadbabaie et al., 2012). Bayesian learning over time on a network generates inference problems about what neighbors know and knew, what their neighbors know

Parameter	Impact on Rebel Success
Persuasiveness of Rebels' Initial Message (B)	↑
Error Rate (ϵ)	↓
Government's Offer (γ)	↓
Marginal Impact of One Informant (M_0)	↑

Table 3: Overview of non-network comparative statics.

and knew, and so on which not only requires carefully accounting for the whole network structure but also makes aggregating the information computationally intensive and often intractable. Models allowing agents to be Bayesian often capture a highly restrictive environment in which all agents have a single decision to make in sequence and learn about all previous agents' decisions (Banerjee, 1992; Bikhchandani, Hirshleifer and Welch, 1992; Smith and Sørensen, 2000) or learn about a random sample of all previous agents (Banerjee and Fudenberg, 2004).[17] More common is to assume agents use simpler rules-of-thumb to aggregate new information received through a network (DeGroot, 1974; Ellison and Fudenberg, 1995; Bala and Goyal, 1998, 2001; DeMarzo, Zwiebel and Vayanos, 2003; Acemoglu, Ozdaglar and ParandehGheibi, 2010; Golub and Jackson, 2012).

We follow the latter approach. By defining links in the network as trusted interpersonal relationships, we assume that a person regards a message received from one of these contacts as credible. For simplicity, we assume that the person's mind is changed by it– she updates her estimation of the rebels based on this rumor. Note that an alternate interpretation is compatible with our setup: it could be that a person is not personally convinced by the message, but instead treats it as a credible signal of what others are thinking. Rumors result in people updating their estimate of the popularity of the rebels, which would have analogous consequences for coordinating on secret-keeping and generate the same comparative statics on the network parameters. In the article text, we present evidence from interviews with civilians and former rebels which suggest that peoples' estimates of others' support for the rebels were quite impressionistic.

The key features of information spread and changed beliefs that drive our results are that the rumors that may persuade a person to keep secrets reach them through network links, and the farther the rumors must travel through the network, the less persuasive/ effective/ clear they are. Any rumor-spreading and belief-updating process that includes these two features will result in networks with low fragmentation and short paths being more favorable to rebel viability than networks with higher fragmentation and longer paths.

1.12 Note on Form of Coordination Problem

The coordination problem captured by the model differs from standard collective action problems. In contrast with traditional free-riding problems, here it can be that just a few informants could do significant damage to the aspiring rebels' probability of success. The more others a civilian expects to keep the rebels' secrets, the *greater* the incentive she has to keep them too. If instead she expects many others to inform, then she has an incentive to join them and inform as well since the marginal damage will be low and she will receive the government's payment.

The herding nature of the coordination problem arises from the specific, peculiar context of early rebellion. Due to their fragility, the clandestine rebels need to remain off the radar

[17] For instance, we could use a semi-Bayesian approach in which players have priors over the benefits from the rebels, regard the message from the rebel as a noisy signal, and use Bayes' rule to update their posterior with the first message to reach them. This setup would generate the same comparative statics on the network, though would be more cumbersome.

of the government if they are to become viable. Single pieces of information– reports on their movements, or even confirmation that they exist– can be devastating.

Of course later phases of rebellion can feature more classical coordination problems. For instance, high-risk, public civilian mobilization for insurrection can exhibit standard collective action problems (Lichbach, 1998; Petersen, 2001; Popkin, 1979; Wood, 2003). However, we argue that groups cannot reach the point of recruiting or mobilizing further without first surpassing an initial threshold of viability. The coordination problem we present can be thought of as a necessary hurdle; once cleared, groups can face additional, different coordination problems.

Our model admittedly simplifies what is probably a more complicated decision on the part of each civilian. We model the choice as "how much damage will I do to the rebels by informing" versus "how much will I receive from the government," where the first part of the choice depends on expectations about how many others will inform. However, because early secrecy is necessary for all later phases of rebellion– recruitment, arming, success in conflict– any decision that accounts for the ultimate probability of rebel success must account for the likelihood of widespread secret-keeping. For instance, a civilian may also in reality account for the likelihood that rebels would win and punish them for informing. Even this accounting hinges on whether the rebels cross the first hurdle, which hinges on expectations of the number of fellow civilians who keep secrets.[18]

In short, we argue that the coordination problem captured by our model is quite general to the early process of rebel groups becoming viable, and resolving it is necessary, though certainly not sufficient, for ultimate rebel success.

1.13 Note on Endogenizing Choice of Launch Location

Our model suggests that rebel groups that launch in ethnically homogeneous areas have a better chance of becoming viable, and hence ultimately succeeding, than those that launch in ethnically heterogeneous areas. An implication is that, all else equal, rebels would do best to relocate to and launch in homogeneous areas.

However, all else is not equal. In reality, not all locations are available to an aspiring rebel group. One key constraint is the rebels' ability to insert their message into the nearby civilians' network. Rebel groups do well in homogeneous areas *given that they can get their message into the local network.* We point out in the article text that this is easiest when a rebel has a personal tie to someone in the area. Such a person may be more amenable to the rebels' cause and more likely to spread rumors championing them. If rebels travel to a brand new location chosen because it is ethnically homogeneous but have no personal contact there, they may find it difficult to get their framed, compelling message to spread.

In the model with the launch location exogenously given, rebels relay their message to a seed among the civilians, i^{seed}, who forms a judgment, B, about the rebels based on this

[18] Moreover, interviews with civilians reveal no evidence that they were concerned by the prospect of future punishment by rebels if they were to be ultimately successful. While later phases of rebellion– recruitment and fundraising, for instance– may drive civilians to make this calculation, it does not appear to have been a factor in civilians' decisions in either of the two cases of initial rebel *launch* presented in the article text.

message. If the initial assessment B is sufficiently poor, the rumor that spreads will not result in civilians choosing to keep secrets, even in homogeneous areas. Endogenizing the choice of launch location would need to account for the dependence of the seed's assessment B on this choice. Doing so would reveal the importance of a location expected to generate a high B. Since in practice this likely depends on preexisting close personal ties, the feasible set of launch locations may be very small.[19]

Indeed, the empirical record supports the constrained choice that aspiring rebels face. If the choice of launch location were endogenous and without constraint, rebels should only launch in ethnically homogeneous areas. Instead, as we show with the analyses of rebel *launch* below, rebels launch in both homogeneous and heterogeneous areas (Table 5). We further show with the analyses of rebel *viability* below that given rebel groups have launched, those that do so in homogeneous areas are indeed more likely to become viable (Table 7). These empirical regularities support both the rebels' constrained choice of launch location and the value of being able to launch in homogeneous areas.

[19] Other constraints may bear on the choice of launch location as well. Although rebels need to surpass initial viability, they will also eventually need to become a powerful fighting force. Expectations of recruitment pools, funding sources, strategic hiding places, and the like may also lead some to gamble on launching in a heterogeneous area if the other factors that will matter later on are favorable.

2 Rebellion in Uganda since 1986

In January 1986, the National Resistance Movement (NRM) seized the central government of Uganda. Prior to 1986, Uganda had suffered from 15 years of relative chaos and state decay: eight years under the notorious Idi Amin (1971-1979), followed by a civil war known as the Bush War. An estimated 800,000 Ugandans were killed as a result of political violence between 1971 when Amin seized power and the end of the civil war in 1985 (Tripp 2004, 4). The Bush War also had a major impact on the Ugandan economy, leading to high inflation and economic decline; in 1985, income per capita was at just 59% of its 1971 level (Kiyaga-Nsubuga, 2004, 89). One Ugandan scholar writes: "By the time Obote fell in 1985, Uganda could no longer be described as a state. No one military or political organization commanded the legitimate use of violence in the country" (Kasozi, 1994, 193).

As described in the article text, as the new NRM government turned towards rebuilding the Ugandan state, it faced numerous insurgencies; in its first two decades in power, 16 distinct groups *launched*, which we define as occurring when a group builds a nascent organization with a command structure and commits (or explicitly plans to commit) at least one act of violence against the Ugandan state (e.g. military, police, or other government targets).[20] The command structures of each of these organizations was clearly distinct, and while some of the rebel groups occasionally communicated with one another and weighed coordinating arms shipments or attacks, actual coordination was extremely infrequent. No two rebel groups operated in the same region of Uganda at the same time.

Of these 16 rebel groups, only four became viable. Recall from the article that we define *viable* groups as those that at least minimally challenged the authority of the central government, which we operationalized as groups that maintained an operational base on the target country's soil with at least 100 individuals for at least three months.[21]

2.1 Analysis of Rebel Launch

While the outcome at the center of this article is whether or not rebels become viable, we argue that there are no systematic relationships regarding where rebel groups initially launch in Uganda that then influence which groups become viable – or at least confound the relationship between ethnic networks and rebel viability that we highlight in the main text. To substantiate this claim, we present an analysis here of the correlates of initial rebel group formation in Uganda since 1986. Recent findings in the quantitative literature on conflict onset identify numerous variables that are correlated with the onset of violent

[20]Only one of these groups, the LRA, is captured in the Correlates of War dataset; seven are captured in the finer-grained UCDP/PRIO Armed Conflict Dataset. Arriving at this number of distinct groups and ensuring that each group met the criteria described in the introduction for constituting a nascent rebel group entailed triangulating among documents from Uganda's Amnesty Commission, a complete set of Ugandan newspaper articles on armed conflict from this period, and interviews conducted by one of the authors with sources from military intelligence, former rebels, government officials, and civilians from communities where rebel groups formed.

[21]The analysis below is not sensitive to these precise thresholds.

organized conflict; we examine whether those relationships apply to initial rebel group launch in Uganda – if they do, it is important to control for them in the analysis of whether already-formed groups become viable. Specifically, we use logistic regression analysis of the location of rebel group launch at the county level, measured as whether a rebel group's first attack occurred in a county.

2.1.1 Measurement

We use cross-sectional data because the best data available for most covariates change little or not at all over time. For example, geographic features and distances do not change (except perhaps for forest cover, for which unfortunately only measures for one year is available), and Uganda's most high quality, high resolution demographic and poverty data (via the census) are only available in two years: 1991 and 2002. Because all but one Ugandan rebel group formed prior to 2000, adding the 2002 data would not add much information to the analysis.[22]

We note that the geographic dispersion of these rebel groups (shown in the article text) suggests that sparks of rebellion do not occur only on the periphery, as leading theories might lead us to expect – they occur all over the country, even close to the capital city, Kampala. The Moran's I index of these points (which are positioned at the centroid of each county where groups first committed violence) is -0.01 and is far from statistical significance, indicating that these points are neither clustered nor dispersed; there is no spatial autocorrelation among them.

We measure covariates at the county level for several practical reasons. Most importantly, the county is the lowest administrative unit for which we could identify locations of initial rebel group attacks. Discerning and verifying these locations involved collecting and triangulating information from interviews (with former rebel leaders, government officials and civilians in the areas of Uganda where these groups formed), data from the Government of Uganda (particularly Chieftancy of Military Intelligence responses to information requests and data from the Amnesty Commission), and newspaper articles. Second, unlike districts in Uganda, the number and boundaries of counties have remained stable over time–only one new county has been created in Uganda since the 1990s, while the number of districts has expanded from 38 in 1990 to over 100 today. There are 163 counties in Uganda, with a median size of 1,045.5 square kilometers per county, which is less than one-third the size of Rhode Island.

We used the following data and measurements, which are also summarized below in Table 4. The literature on civil conflict onset suggests the potential importance of several variables to the early phases of rebellion, some of which can be set aside here because they are not present in the Ugandan context and thus their relative importance cannot be empirically evaluated. For example, Uganda does not have high-value, lootable resources (Collier and Hoeffler, 2004; Lujala, 2010), and none of the rebel groups there benefited from significant material resources in the initial stages (Weinstein, 2007). This literature also

[22] Furthermore, ethnic demographic patterns change little during this period; for example, county-level ethno-linguistic fractionalization (ELF) scores in 1991 and 2002 are correlated at 0.9.

tends to distinguish between proxies for motivation to rebel versus feasibility of rebellion. In the family of motivation-focused explanations for conflict onset, we use measures of local ethnic demography and of exclusion from the central government, as follows:

- We measured *exclusion* from the central government using Stefan Lindemann's data for the ethnic composition of Uganda's cabinet in 1988. First, we developed an exclusion score for each ethnic group by subtracting the ethnic group's share of Uganda's total population from its share of the cabinet (Lindemann, 2011, 397-400).[23] We then developed exclusion scores for each county according to a weighted average based on the ethnic composition of each county. Negative scores indicate"under-representation" in the cabinet, meaning that on average, the ethnic groups in a given county are over-represented and positive scores indicate "over-representation." The distribution of this measure, however, is right skewed because the Baganda are quite overrepresented; the median exclusion score for ethnic groups is -.75, and while the lower bound of the range is -3.4, the Baganda (at the upper bound) score 17.4. To deal with this issue, we also code a trichotomous version of ethnic cabinet representation that takes 0 if the ethnic group is under-represented, with a score below -1, a 1 if it is not substantially over or under-represented, with scores between -1 and 1, and 2 if it scores above 1. Then, we again associate these ethnic cabinet representation scores with counties using a weighted average based on each county's ethnic composition. For both of these measures of exclusion, most existing theories posit that more under-represented areas will be more likely to rebel.

- To measure the *ethnic demography* of a county, we construct two variables. The first is the ethno-linguistic fractionalization (ELF) score, which measures the probability that two, randomly-sampled individuals from a county would be from different ethno-linguistic groups. The data we use to compile this measure comes from Uganda's 1991 census. As an alternative measure to ELF, we construct a second measure of ethnic demography: the share of a county's total population that is comprised by the largest ethno-linguistic group in that county. We construct this second ethnic demography variable because it could in principle measure different underlying ethno-political dynamics than the first. However, for Uganda's counties, this measure is almost perfectly negatively correlated with ELF. Thus, we show results below using the ELF measure below, but the results are almost identical (with opposite signs on the coefficients) when using the latter measure in place of ELF. Contrary to our theory, which expects no relationship between ELF and rebel launch (see above in Section

[23] We use this data source rather than the prominent Ethnic Power Relations (EPR) dataset since this data allows us to use fine-grained, direct measures of political exclusion for all ethnic groups in Uganda with over 1% of the population. EPR codes a subset of these groups; those which their coders deemed politically relevant (Wimmer, Cederman and Min, 2009). Because of the groups not deemed politically relevant were leaders of incipient rebel groups in Uganda, we follow Fearon (2003) in considering all ethnic groups that comprise over 1% of the population relevant. Except for the Baganda (who were 18% of the population,) as of the 1991 census, no ethnic group in Uganda was more than 10% of the population.

1.13), most existing theories would predict that more homogeneous areas will be more likely to experience rebel initiation.

To capture the feasibility of rebellion, we measured variation in terrain, distance from international borders, local development, and local state capacity:

- To capture variation in *terrain* that may facilitate rebel viability, we first generated an elevation variable that measures the difference between the highest and lowest point in a county, using data compiled by the U.S. Geological Survey. We also constructed a measure of forest cover, measuring the percentage of each county that has natural (non-agricultural) vegetation, such as trees and shrubs, using data from the U.N. Food and Agriculture Organization (FAO)'s Africover project. As an additional measure, and because population density is plausibly lower in thickly forested or otherwise "rough" areas that may be favorable for rebellion, we also compiled a population density variable (logged) using the complete 1991 census, measuring the number of people per square kilometer. The expected relationship, based on existing theories, is that areas that are more mountainous (higher difference in elevation), more forested, or less populous will be more likely to experience rebel initiation.[24]

- To examine the importance of access to *international borders* with states that may be advantageous for rebels who can use the cross-border area as a fallback, we measure the distance (as a straight line, logged) from the centroid of the county to the closest international border using ArcGIS. As a second measure, in order to capture that possibility that what matters to nascent rebels is having direct access to an international border, rather than relative proximity, we create a binary variable to indicate whether or not a county is on an international border.[25] The expected relationship, based on existing theories, is that the closer a county is to an international border (or if it is contiguous with a border), the more likely it is to experience rebel initiation.

- To measure the extent of *local development*, we use a county-level 1992 measure of the percentage of a county's population living below the poverty line, obtained from the Ugandan Bureau of Statistics report,"Where are the Poor? Mapping Patterns of Well-Being in Uganda" (Emwanu et al., 2003).[26] We also use a measure of the poverty

[24]These measures are not highly correlated: The correlation between forest cover and elevation range (distance between highest and lowest point) is 0.41; the correlation between forest cover and maximum elevation is 0.37; the correlation between population density and elevation range is -0.13; the correlation between forest and population density is -0.44.

[25]This measure is somewhat correlated with the distance from the capital measure; the correlation between the two measures is -0.66. There were weakly-governed areas across all of Uganda's borders during this period, so for simplicity we treated all borders equally. For example, in the late 1980s and early 1990s, eastern Ugandan rebel groups viewed western Kenya as a fall back option. We only considered land borders we did not count lakes as international borders since large bodies of water do not offer rebels an appealing fall back option, and fieldwork did not indicate of any cases of rebel groups hiding in, or smuggling weapons via, boats.

[26]The poverty line used is the official one adopted by the Government of Uganda, according to the methodology in Ravallion (1994). For basic information about Uganda's poverty line see Emwanu et al. (2003, 13)

gap, which measures how far (as a percentage) below the poverty line is the average consumption of poor people (those living below the poverty line), providing insight into how deep poverty is in a given county. However, this measure is almost perfectly correlated (0.96) with the former poverty line measure, and thus the results of the below analyses are quite similar using either measure. We show only results for the poverty line measure in what follows. While there are several theoretical perspectives on why poverty influences conflict onset, all would expect a positive relationship between poverty and rebel initiation. These poverty measures are based on county-level estimates of household consumption information collected in Uganda's 1992 Integrated Household Survey, combined with information from Uganda's 1991 census.[27]

- Following Collier and Hoeffler (2004) and others (e.g. Do and Iyer (2010); Thyne (2006); Walter (2004)), we compiled county-level *literacy* rates as an alternative proxy for living conditions or the opportunity cost of participating in rebellion. The data source is the 1991 Uganda census. The correlation between percentage of people living below the poverty line and literacy is -0.57. Most dominant theories expect that in areas with higher literacy rates, individuals will be less likely to rebel, because of the higher opportunity costs of rebelling implied by a literate, and thus employable, population.

- To measure *state strength*, we create a binary variable capturing whether or not the county was under NRM control during the Bush War. Because the NRM controlled some parts of central and western Uganda during the Bush War and developed local political and intelligence institutions in the areas it controlled, the state may plausibly have had higher capacity to respond to potential threats in those areas. Additionally, because the capital, Kampala, was the institutional center of the state and the military when the first anti-NRM rebel group formed in August 1986, we create a second measure of state capacity: distance from the Kampala (logged). This distance is measured (using ArcGIS) as a straight line from the centroid of each county to the center of Kampala. Most extant theories would expect that the further a county is from Kampala, the more likely it to have a rebel group form there.

[27]This estimation (rather than a direct poverty measure from the household survey) is necessary because the household survey, which is the only one conducted in the early 1990s in Uganda, was designed to be representative only at a higher level (regions). The results of the survey are therefore not representative at the county level (there are four regions and 163 counties in Uganda). However, using Small Area Estimation techniques with information from variables that were directly measured in both the household survey and the census, the Ugandan Bureau of Statistics (in collaboration with the International Livestock Research Institute) generated the estimates of county-level poverty we use here. For more information about how this measure was compiled, see Emwanu et al. (2003, 13), especially Appendix 1. These measures provide the best available estimates of poverty in Uganda's counties in the early 1990s. An issue with this data is that several rebel groups formed prior to 1992 and local consumption is likely highly responsive to violence; unfortunately reliable, local economic welfare data on Uganda in the 1980s does not exist. This measure will capture any increases in poverty that occurred as a result of violence in areas where rebels formed prior to 1992, therefore biasing results in favor of finding a stronger relationship between poverty and rebel formation than the "true" relationship.

To control for possible regional effects, we also include regional dummies in all model specifications. We also cluster standard errors by region.

Table 4: **Summary Statistics, Variables for Rebel Launch Analysis**

Variable	Mean	SD	Range	Description	Data source
formation (DV)	0.10	0.30	{0, 1}	"1" if a rebel group committed first attack in that county, otherwise "0"	Interviews with rebels, government, and civilians; newspaper articles
ELF	0.38	0.25	(.01, .88)	Probability that two randomly-selected people are from different ethnic groups	Uganda census (1991)
largest ethnic group size	.74	.20	(0.22 , 0.99)	Portion of the county's total pop. comprised by largest ethnic group (logged)	Uganda census (1991)
cabinet representation	2.32	7.49	(-2.9, 16)	Weighted average of ethnic groups' cabinet share minus Uganda pop. share	Lindemann (2011)
cabinet rep 2	0.82	0.74	(0.0002, 1.99)	Same as above, except scores are trichotomous: diff. between ethnic cabinet share and pop share score is "0" if diff. is < -1, "1" if > -1 and < 1, "2" if > 1	Lindemann (2011)
elevation (km)	1,731	815	(912, 4,858)	Difference between highest and lowest point (km)	USGS
forest	24.75	25.86	(0, 95.23)	% area comprised of non-agricultural vegetation	UN FAO (2002)
population density (log)	4.55	1.38	(0.07, 9.23)	No. of people per sq km (log)	Uganda census (1991)
distance to border (m) (log)	10.87	0.89	(8.84, 12.13)	Distance in m from county centroid to closest int'l border (log)	Measured using GIS
contiguous with border	0.21	0.41	{0, 1}	Is country on international border	n/a
poverty line	63.46	13.94	(26.1, 93.1)	Percentage of people living below the poverty line in 1992	Uganda Bureau of Statistics (2003)
poverty gap	23.38	9.72	(3, 49.4)	Among those living below poverty line, average consumption (% below pov line) in 1992	Uganda Bureau of Statistics (2003)
literacy	50.68	15.02	(0.003, 88.39)	% literate in 1991	Uganda Bureau of Statistics (2003)
distance to capital (m) (log)	12.04	1.07	(1, 12.89)	Distance in m from centroid of country(ies) to closest int'l border (log)	Calculated using GIS
prior NRM institutions	0.33	0.47	{0, 1}	Was county under NRM's control prior to January 1986	Interview with senior NRM official; newspaper articles

Note: Unit of analysis is Uganda's 163 counties.

2.1.2 Results for Rebel Launch

Table 5 displays the results of logistic regression analyses with robust standard errors clustered by region.[28] Consistent with the claim above that rebel launch does not follow strong patterns that could impact which groups become viable, the main finding of this exercise is that there are few consistent, statistically significant patterns that are consistent with prior theories relating these variables to conflict onset – with the exception that counties that are on an international border are more likely to experience rebel formation. An additional finding, discussed below, is that there is some evidence for a positive relationship between literacy and rebel formation.

Model 1 shows the full model, which includes measures for each of the variables discussed above. We include all measures of the terrain and development variables in the full model since they measure different types of rough terrain, and the measures are not highly correlated. We only include one measure of proximity to international borders, state capacity,

[28]All models include regional dummy variables. None are significant, except northern is significant in model 7 only.

and exclusion in the full model since in these cases the alternative measure captures a highly similar concept. Models 2 and 3 check whether the results of the full model hold when using alternative measures of these three variables. Literacy is positive and highly significant in all models, although its substantive effect is not large; for example, in a simulation using model 3, holding all other covariates at their mean and increasing the literacy variable from its 25th percentile to its 75th percentile raises the likelihood of rebel formation by 4.6 percentage points. Contiguity with an international border is significant in all models and has a larger substantive effect; in a simulation using model 3, holding all covariates at their mean while changing the contiguous border variable from zero to one yields an 14.6 percentage point increase in the predicted probability of a rebellion occurring in that county.

In models 1, 2 and 3, all measures of ethnic exclusion and demography, terrain, and spatially-varying state strength are insignificant in all models, with the exception of the population density's significance in model 3 only. Most of the variables do take the sign anticipated by existing theories: conditional on the other variables, change in elevation is positively related to rebellion launch; population density is negatively related; measures of state strength are negatively related; distance to international borders is negatively related; and poverty is positively related. One exception is percent of forested area, which we would expect to be positively related to rebel formation but comes through here as negatively related – however, the substantive influence of this negative relationship is quite small. The sign on ELF is not consistent across models; this lack of a relationship is consistent with our theory, which posits that ethnic demography does not influence where groups launch, since aspiring rebel leaders have a limited set of options for locations to launch.

Another exception in which a variable does not take the expected sign is for the first measure of ethnic exclusion, for which counties are coded according to a weighted average of their ethnic composition, based on each ethnic group's over or under-representation in the cabinet. Excluded groups take negative values. The coefficient on this covariate is positive, contradicting conventional wisdom that less excluded groups would be less likely to rebel. This result is likely due to the influence of the counties with a large portion of Baganda, who are much more over-represented than any other group is; three of the 16 rebel groups formed in majority-Baganda areas. Comparing model 1 and model 2 makes this clear; the coefficient on exclusion switches signs when substituting the original exclusion measure for the second measure, which dampens the influence of the Baganda's extreme overrepresentation by making extent of ethnic exclusion a categorical (three-category) variable.

Models 4 and 5 address the possibility that poverty, education, or exclusion are influenced by some of the other factors which also have a direct influence on rebel formation, and thus those other factors' relationship with rebel formation would be attenuated or exaggerated in models 1, 2, and 3. For example, it is possible that being closer to the capital, or having been occupied by the NRM during the Bush War, would make it less likely that a county would be poorer, less educated, or underrepresented in the cabinet. This may be the case if we believe, for example, that employment and educational opportunities are positively correlated with proximity to prior presence of NRM local institutions, and if areas more proximate to the capital or that had NRM presence prior to 1986 are better able to pressure the NRM governments into including their representatives in the cabinet. If these conjectures are correct, then models that include, for example, both distance from Kampala and exclusion from the cabinet covariates would bias downward the effect of distance from Kampala on rebel formation. Comparing model 4 with model 2, and model 5 with model 3, respectively, shows that the absolute value of the magnitude of the coefficients of state capacity increases slightly, so these conjectures may be correct.

In model 6, we probe the influence of poverty and literacy, since it is plausible that literacy has a direct influence on both poverty and rebel formation, including it along with poverty may attenuate the effect of the poverty variable. In model 7, we check the possibility that poverty is attenuating the coefficient on literacy. However, in models 6 and 7 the substantive and statistical significance of the variables changes little from prior models.[29]

In sum, the incidence of initial rebel group launch in Uganda appears to be rather idiosyncratic, or at least, its spatial distribution is not consistent with most existing explanations. An exception is that this analysis finds that counties that are contiguous with an international border are more likely to experience rebel group launch, or at least to have a new rebel group launch its first attack there, than those that are not contiguous with a border. Another finding of note, although the relationship is not particularly strong, is that literacy is positively related to rebel group launch. This latter finding contradicts most existing explanations, which consider literacy to be a proxy for development and to indicate individuals' opportunity costs for participating in rebellion. Given that most nascent rebel leaders in Uganda were reasonably well-educated, this suggests that rather than proxying for the opportunity costs of rebellion, higher local levels of literacy may indicate a concentration of people who have higher capacity for taking on the complex, risky enterprise of forming a rebellion.

For the purposes of this paper, the main potential issue that these findings raise is the possibility that patterns affecting where rebels initially launch could influence the subsequent stages when they attempt to become viable – that is, if certain "types" of rebels are more or less likely to form on borders or with a more literate population, then that difference in "type" could be the reason why only some rebels become viable. Because of this problem, in our paired case comparison, we compare cases with roughly similar levels of literacy. Additionally, only one group we compare (FOBA) launched in a county on an international border, and contrary to extant theories, this is the group that failed to become viable.

2.2 Analysis of Rebel Viability

The article establishes a correlation between local ethnic demography of areas where rebels launch and whether or not those groups become viable, and explains that trivariate regressions provide added evidence that this relationship may not be spurious. This section presents these regression analyses in greater detail. Recall from the article that we define viable groups as those that challenged the authority of the central government, which we

[29] For both models 6 and 7, these results remain substantively and statistically similar when including either measure of exclusion.

Table 5: Correlates of Initial Rebel Group Launch in Uganda

	1	2	3	4	5	6	7
Exclusion							
cabinet representation	0.02		0.05				
	(0.20)		(0.15)				
cabinet rep. 2		-0.40					
		(0.78)					
Ethnic Demography							
ELF	0.24	0.21	0.58	0.37	-.023	-0.23	0.47
	(1.86)	(1.92)	(1.60)	(1.60)	(1.43)	(1.43)	(1.29)
Terrain							
elevation (km)	0.54	0.53	0.46	0.42	0.32	0.32	0.45
	(0.71)	(0.58)	(0.75)	(0.65)	(0.67)	(0.67)	(0.70)
forest	-0.02	-0.02	-0.01	-0.02	-0.02**	-0.02**	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)
pop. density (log)	-0.25	-0.24	-0.27**	-0.13	-0.13	-0.13	-0.30**
	(0.19)	(0.24)	(0.09)	(0.21)	(0.14)	(0.10)	(0.10)
International Borders							
distance to border (m)(log)	-0.47	-0.63		-0.34			
	(0.36)	(0.45)		(0.22)			
contiguous w border			1.55**		1.37**	1.36**	1.65**
			(0.59)		(0.51)	(0.46)	(0.54)
Development							
poverty line	0.03	0.03	0.02			0.00	
	(0.04)	(0.04)	(0.04)			(0.03)	
literacy	0.06*	0.06^{+}	0.06**				0.06**
	(0.03)	(0.03)	(0.02)				(0.02)
State Capacity							
prior NRM	-0.58	-0.69		-0.71			
	(0.50)	(0.81)		(0.49)			
dist. to capital (m) (log)			-0.06		-0.10	-0.11	-0.02
			(0.10)		(0.04)	(0.11)	(0.04)
region dummies	X	X	X	X	X	X	X
N	161	161	161	161	161	161	161

DV is a dichotomous measure of whether a rebel group committed their first attack in the county. All models are estimated using logistic regression with robust standard errors, clustered by region (central, northern, eastern, or western), reported in parentheses.

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

operationalized as groups that maintained an operational base on the target country's soil with at least 100 individuals for at least three months.[30]

2.2.1 Measurement

Measures used for this analysis are the same as those for the rebel launch analysis above. Table 6 below displays summary statistics for the variables.

The unit of this analysis is the rebel group. However, many of the variables we seek to measure across rebel groups for this analysis are distributed across space, such as the ethnic demography of the area in which rebels operated in the initial stages, defined here as where nascent rebels went during the initial stages of their emergence for the purpose of (1) attempting to establish a base or (2) interacting with and seeking material support (money, food, or recruits) from civilians. The initial stages are conceptualized here as the period between the inception of the insurgent group, when planning begins, until a group achieves viability or is defeated, ceasing to attack Uganda – typically following a negotiated settlement or following the capture, exile, or death of one or more top leaders. These areas were determined by Ugandan newspaper articles on these rebellions, and because these were often highly incomplete, the fieldwork of one of the authors, which is described in the main article text. The initial activities of four rebel groups (ADF, HSM, UFDF, CAMP) were clearly associated with a single county, so geographic variables associated with those rebel groups correspond to that county. However, some rebel groups (UPA, LRA, NFA, WNBF, UNRF II, NALU, FOBA) initially spanned two, three, or in one case, four contiguous counties. For those groups, the variables were measured using data from all relevant counties. For example, the ethnic demography and population density measures for FOBA were derived using census data for all individuals living in both Tororo and Samia-Bugwe counties, and the highest elevation reported is the highest point present in either county. For five rebel groups (NOM, UPDA, PRA, NDA and UDA), it was not possible to reliably discern precisely in which county or counties they operated in during the initial stages. In those cases, it was clear that the group likely operated in at least one of three to five counties and thus for all spatially-measured variables, the average values over those three to five counties was used.

To measure ethnic demography, we use 1991 Ugandan census data to construct an ethnolinguistic fractionalization (ELF) score of the county(ies) where each rebel group operated in the initial stages. Based on the logic laid out in the main article, higher ELF scores in a given area should indicate more fragmented information networks among people living there. None of the results presented here are substantially changed when using an alternative measure of ethnic demography, the percentage of the total population of the area comprised by the largest ethnic group.

The use of 1991 census data is necessary because reliable census data at the county level is not available prior to 1991; however, this introduces potential measurement error for the nine rebellions we examine here that began before 1991. Violent conflict often impels migration,

[30]The analysis here is not sensitive to these precise thresholds.

plausibly making an area increasingly more homogenous in the period after conflict begins. This issue could bias our results in the direction that favors our hypothesis. However, this issue affects areas with different baseline demographic patterns similarly, biasing the measurement of all areas in the same direction and thus limiting the biases of inferences about the effects of differences in ethnic demography across areas on rebel viability – unless we believe that the groups that became viable were more successful at ethnic cleansing. It is difficult to rule out this possibility; however, six of the nine insurgencies that began prior to 1991 took place in eastern or northern Uganda, where there is ample historical evidence that ethnic demographic patterns in these regions have persisted for decades – long before these rebellions began (see the main article for a discussion of this evidence for the parts of eastern Uganda where FOBA and UPA formed.) To ensure that any potential biases affecting the remaining three insurgencies – NOM, NDA and NALU – are not driving the results presented below, we re-ran the analysis without these three observations, and find that they do not substantially alter the results; ELF retains significance throughout all models at least at the 15% level.

Table 6: **Summary Statistics, Variables for Rebel Viability Analysis**

Variable	Mean	SD	Range	Description	Data source
viability (DV)	0.25	0.45	{0, 1}	"1" if became viable, otherwise "0"	Interviews with rebels, government, and civilians; newspaper articles
ELF	0.40	0.25	(.02, .76)	Probability that two randomly-selected people are from different ethnic groups	Uganda census (1991)
largest ethnic group size	.70	.21	(0.33 , 0.99)	Portion of the county's total pop. comprised by largest ethnic group (logged)	Uganda census (1991)
exclusion	1.37	5.15	(-2.72, 12.61)	Weighted average of ethnic groups' cabinet share minus Uganda pop. share	Lindemann (2011)
elevation (km)	1,900	1,113	(1015, 1,196)	Difference between highest and lowest point (m)	USGS
forest	26.59	17.03	(1.24, 64.23)	% area comprised of non-agricultural vegetation	UN FAO (2002)
population density (log)	4.40	0.64	(2.95, 5.23)	No. of people per sq km (log)	Uganda census (1991)
distance to border (m) (log)	10.72	1.11	(9.14, 12)	Distance in m from county centroid to closest int'l border (log)	Calculated using GIS
contiguous with border	0.5	0.52	{0, 1}	Is country on international border	n/a
poverty line	66.2	12.7	(43.4, 91.40)	Percentage of people living below the poverty line in 1992	Uganda Bureau of Statistics (2003)
poverty gap	25.6	8.34	(13.7, 47.43)	Among those living below poverty line, average consumption (% below pov line) in 1992	Uganda Bureau of Statistics (2003)
literacy	48.67	9.55	(29.56, 66.77)	% literate in 1991	Uganda Bureau of Statistics (2003)
distance to capital (m) (log)	12.18	0.84	(10.04, 12.86)	Distance in m from centroid of country(ies) to Kampala (log)	Calculated using GIS
state strength	1.25	1.29	{0, 1}	Was area under NRM's control prior to January 1986	Interview with senior NRM official; newspaper articles

Note: Unit of analysis is the rebel group. Geographic variables correspond to the county(ies) where rebel groups initially operated.

2.2.2 Results for Rebel Viability

Regression analysis of these variables using a linear probability model provides useful insight. Because of the small number of observations, we run trivariate models, with ELF and one of the control variables in each regression. The results of the regression analyses are summarized below in Table 7.

Table 7: Correlates of Initial Rebel Group Viability in Uganda

	1	2	3	4	5	6	7	8	9
ELF	-.873$^+$	- .898$^+$	- .899$^+$	-.826$^+$	-.870$^+$	-.873$^+$	-.894$^+$	-.806$^+$	-.716
	(.418)	(.489)	(.445)	(.447)	(.429)	(.436)	(.440)	(.451)	(.431)
elevation		.000							
		(.000)							
forest			-.002						
			(-.007)						
pop density				-.066					
				(.198)					
dist. to border					.003				
					(.088)				
literacy						.000			
						(.005)			
poverty line							-.001		
							(.008)		
state capacity								-.083	
								(.121)	
exclusion									-.013
									(.010)
N	16	16	16	16	16	16	16	16	16

DV is a dichotomous measure of rebel viability.

All models are estimated using OLS with robust standard errors, reported in parentheses.

Results are not substantively changed when using alternative measures of ethnic demography, poverty, or state capacity. Alternative measures are shown in the table of summary statistics above.

$^+$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

We present OLS results above because of the ease of substantive interpretation. In all models except one, the coefficient on ELF is significant at the 90% level, whereas none of the other covariates are statistically significant. (The sole model where ELF is not significant at the 90% level, model 9, is significant at the 85% level; the p-value is 0.12). The marginal effects resulting from logit models with the same covariates indicate nearly identical substantive results. These results are also not driven by any single group, nor are they substantially altered when replacing ELF with a measure of the percentage of the population comprised by the largest ethnic group in the area.

Naturally, the results of any regression analysis with so few observations must be interpreted with great caution. However, the fact that ELF is statistically significant in almost all models and withstands several sensitivity tests provides useful support that is consistent with the central hypothesis of the article. Furthermore, in light of the substantive and statistical non-significance of the control variables, it is unlikely that the correlation between ethnic homogeneity and viability is spurious.

3 Case study selection

As noted in the paper, the case studies were selected with the aim of isolating the variable of interest: local ethnic demography – which, as we show in the article, corresponds to differences in trusted network structure – while "holding constant" other, potentially-relevant factors drawn from dominant theories of civil war onset. The factors "held constant," or at least that varied in the direction other than that expected by existing theory, are summarized below.

	UPA *Teso region of eastern Uganda*	**FOBA** *Bukedi region of eastern Uganda*
Traditional leadership	Decentralized	Decentralized
Political Exclusion	Low	Low
Terrain	Flat, semi-arid savannah	Somewhat hilly, some jungle and nearby forests
Proximity to international borders	No international border, 120 km to western Kenya	Borders western Kenya
Economic well being (Poverty data from 1992 household survey; literacy data from 1991 census)	77.4% living below poverty line; 51.5% literate	57.7% living below poverty line; 53% literate
State strength *year of rebel launch*	Low *1987*	Low *1987*
Ethno-linguistically homogeneous?	*Yes* *1 language in Teso* *Largest ethnic group is 85% of population*	*No* *3 languages in Bukedi* *Largest ethnic group is 37.5% of population*
Did the rebellion become viable?	***Yes***	***No***

Figure 4: *Summary of Paired Comparison, Case Study Rebellions.*

With respect to the first two factors – type of local traditional leadership and political exclusion – the commonalities among these cases are straightforward: both rebel groups launched in areas in which traditional authority structures were decentralized, and as described in the paper, neither were markedly under represented in the national government in the late 1980s. Unlike areas of central and western Uganda that have been arranged into centralized, hierarchical kingdoms since the pre-colonial period, the areas of eastern Uganda studied in this paper have long been characterized by highly localized leadership structures: traditional leaders for a given kinship group would adjudicate disputes among clans, and while these clan leaders would sometimes coordinate, none was not subject to any overall authority.

Regarding terrain and borders, the variation that exists among these cases goes in the direction opposite of that anticipated by existing theories. Uganda sits astride the equator, much of the country is semi-forested and hilly, with most of the geographic variation being simply the extent of the forest cover and hills. While Uganda is bordered by two mountainous areas on its eastern and western flanks, neither is located close to, or played a significant operational role in, the initial stages of either case study rebellion. Some areas of Uganda are flatter and less thickly forested than others; the UPA – which became viable – launched in the area with the least favorable terrain, since it is the most flat and has the least forest cover.[31] Regarding international borders, FOBA, the group that did not become viable, launched closer to an international border than the UPA, which did become viable.

There are also no major, discernable differences in state capacity or poverty in these areas. The key determinant of economic welfare and state capacity in Uganda in the decade leading up to 1986 – warfare – did not directly affect Teso or Bukedi. The Bush War of the early 1980s, which led to the NRM's ascendance to central power, was fought exclusively in central and western Uganda. Thus, the eastern region studied in this paper escaped the direct effects of war, particularly war-related deaths, destruction of infrastructure and population displacement.

While reliable measures of economic well-being for these areas do not exist for the years immediately prior to these rebellions, measures from the 1991 census provide some clues. As current theories predict, the area in which groups became viable were somewhat poorer than those where groups did not become viable: the percentage of individuals living below the poverty line was 77.4 and 57.7, respectively, for the UPA and FOBA. However, because this data was collected about three years after the start of these rebellions, it is likely that events occurring in the initial stages of these rebellions drive at least some of this variation. Literacy rates may be a better proxy for economic well-being since they are stickier, and indeed respective literacy rates for these areas suggest that areas that spawned viable groups were similarly literate, or more literate, than areas that did not: literacy rates for the area where the rebel groups formed respectively are 51.2 and 53 for UPA and FOBA. Thus, while three years after rebellion began, the areas where rebel groups became viable were somewhat poorer than the areas where they did not become viable, literacy rates were similar across these areas.

[31] In fact, one UPA leader remarked about the UPA disadvantage relative to FOBA with respect to terrain, stating that "the forest [in Bukedi where FOBA launched] is ideal for war. During our rebellion, I would look there with envy."

Regarding state capacity, as a part of their war-fighting efforts in central and much of western Uganda, the NRA had developed significant local institutional capacity. In particular, they had developed extensive local government structures that assisted the NRM rebel group in obtaining information about government troops. These institutions were thus already in place and were absorbed into the national government structure when the NRM overtook Kampala in early 1986. But by the time the rebellions began in northern and eastern Uganda in 1987, the NRM government had only recently begun setting up similar local governance institutions throughout northern and eastern Uganda, appointing Special District Administrators to organize those local governments and hold local elections. Thus, throughout eastern Uganda, the central government had very little institutional reach, and thus very low capacity to detect nascent rebels.

In conclusion, to the extent possible, these cases are paired by factors other than ethnic demography that we may expect to influence rebel viability. There appears to be no compelling reason to believe that any of these factors are strongly associated with rebel viability in these cases in the direction anticipated by existing theories.

References

Acemoglu, Daron, Asuman Ozdaglar and Ali ParandehGheibi. 2010. "Spread of (mis) information in social networks." *Games and Economic Behavior* 70(2):194–227.

Bala, Venkatesh and Sanjeev Goyal. 1998. "Learning from neighbours." *The review of economic studies* 65(3):595–621.

Bala, Venkatesh and Sanjeev Goyal. 2001. "Conformism and diversity under social learning." *Economic theory* 17(1):101–120.

Banerjee, Abhijit and Drew Fudenberg. 2004. "Word-of-mouth learning." *Games and Economic Behavior* 46(1):1–22.

Banerjee, Abhijit V. 1992. "A simple model of herd behavior." *The Quarterly Journal of Economics* pp. 797–817.

Bikhchandani, Sushil, David Hirshleifer and Ivo Welch. 1992. "A theory of fads, fashion, custom, and cultural change as informational cascades." *Journal of political Economy* pp. 992–1026.

Bramoullé, Yann and Rachel Kranton. 2007. "Public Goods in Networks." *Journal of Economic Theory* 135(1):478–494.

Cederman, Lars-Erik, Nils Weidmann and Nils Petter Gleditsch. 2011. "Horizontal Inequalities and the Ethno-Nationalist War: A Global Comparison." *American Political Science Review* 105(3):478–495.

Chwe, M.S.Y. 2000. "Communication and Coordination in Social Networks." *Review of Economic Studies* 67(1):1–16.

Collier, Paul and Anke Hoeffler. 2004. "Greed and Grievance in Civil War." *Oxford Economic Papers* 56(4):563–595.

DeGroot, Morris H. 1974. "Reaching a consensus." *Journal of the American Statistical Association* 69(345):118–121.

DeMarzo, Peter M, Jeffrey Zwiebel and Dimitri Vayanos. 2003. "Persuasion bias, social influence, and unidimensional opinions." *Quarterly Journal of Economics* 118(3):909968.

Dixit, A.K. 2004. *Lawlessness and Economics: Alternative Modes of Governance*. Princeton Univ Pr.

Do, Quy-Toan and Lakshmi Iyer. 2010. "Geography, poverty and conflict in Nepal." *Journal of Peace Research* 47(6):735–748.

Ellison, Glenn and Drew Fudenberg. 1995. "Word-of-mouth communication and social learning." *The Quarterly Journal of Economics* pp. 93–125.

Emwanu, Thomas, Paul Okiira Okwi, Johannes G. Hoogeveen and Patti Kristjanson. 2003. *Where are the Poor? Mapping Patterns of Well-Being in Uganda: 1992 & 1999*. Nairobi, Kenya: The Regal Press Kenya, Ltd.

Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth* 8(2):195–222.

Finkel, Evgeny. 2015. "The Phoenix Effect of State Repression: Jewish Resistance during the Holocaust." *American Political Science Review* 109(2):339–353.

Golub, Benjamin and Matthew O Jackson. 2012. "How Homophily Affects the Speed of Learning and Best Response Dynamics." *Quarterly Journal of Economics* 127(3):1287–1338.

Gould, Roger V. 1991. "Multiple networks and mobilization in the Paris Commune, 1871." *American Sociological Review* pp. 716–729.

Gould, Roger V. 1993. "Collective Action and Network Structure." *American Sociological Review* 58(2):182–196.

Gubler, Joshua R and Joel Sawat Selway. 2012. "Horizontal inequality, crosscutting cleavages, and civil war." *Journal of Conflict Resolution* 56(2):206–232.

Harsanyi, John C. 1995. "A new theory of equilibrium selection for games with complete information." *Games and Economic Behavior* 8(1):91–122.

Jackson, M.O. and A. Wolinsky. 1996. "A Strategic Model of Social and Economic Networks." *Journal of Economic Theory* 71(1):44–74.

Jadbabaie, Ali, Pooya Molavi, Alvaro Sandroni and Alireza Tahbaz-Salehi. 2012. "Non-Bayesian social learning." *Games and Economic Behavior* 76(1):210–225.

Kasozi, A. B. K. 1994. *The Social Origins of Violence in Uganda, 1964-1985*. Montreal: McGill-Queen's University Press.

Kiyaga-Nsubuga, John. 2004. Uganda: The Politics of Consolidation under Museveni's Regime, 1996-2003. In *Durable Peace: Challenges for Peacebuilding in Africa*, ed. Taisier M. Ali and Robert O. Matthews. Toronto: Toronto University Press.

Laitin, David. 2004. "Ethnic Unmixing and Civil War." *Security Studies* 13(4):350–365.

Larson, Jennifer M. 2016. "Interethnic Conflict and the Potential Dangers of Cross-Group Ties." *Journal of Peace Research* 53(3):459–471.

Larson, Jennifer M. 2017. "Networks and Interethnic Cooperation." *Journal of Politics* 79(2):546–559.

Lichbach, Mark. 1998. *The Rebel's Dilemma*. Ann Arbor: University of Michigan Press.

Lindemann, Stefan. 2011. "Just Another Change of the Guard? Broad-Based Politics and Civil War in Museveni's Uganda." *African Affairs* 110(440):387–416.

Lippert, Steffen and Giancarlo Spagnolo. 2011. "Networks of Relations and Word-of-Mouth Communication." *Games and Economic Behavior* 72(1):202–217.

Lujala, Paivi. 2010. "The Spoils of Nature: Armed Civil Conflict and Rebel Access to Natural Resources." *Journal of Peace Research* 47(1):15–28.

Parkinson, Sarah Elizabeth. 2013. "Organizing Rebellion: Rethinking High-Risk Mobilization and Social Networks in War." *American Political Science Review* 107(3):418–432.

Patty, John W and Elizabeth Maggie Penn. 2014. "Sequential Decision Making and Information Aggregation in Small Networks." *Political Science Research and Methods* 2(02):243–271.

Petersen, Roger. 2001. *Resistance and Rebellion: Lessons from Eastern Europe*. Cambridge, UK: Cambridge University Press.

Popkin, Samuel. 1979. *The Rational Peasant: The Political Economy of Rural Society in Vietnam*. Berkeley: University of California Press.

Ravallion, M. 1994. *Poverty Comparisons*. Switzerland: Hardwood Academic Publishers.

Salehyan, Idean. 2010. "The Delegation of War to Rebel Organizations." *Journal of Conflict Resolution* 54(3):493–515.

Salehyan, Idean, David Siroky and Reed M. Wood. 2014. "External Sponsorship and Civilian Abuse: A Principal-Agent Analysis of Wartime Atrocities." *International Organization* 68(3):633–661.

Siegel, David A. 2009. "Social networks and collective action." *American Journal of Political Science* 53(1):122–138.

Smith, Lones and Peter Sørensen. 2000. "Pathological outcomes of observational learning." *Econometrica* 68(2):371–398.

Staniland, Paul. 2012. "Organizing Insurgency: Networks, Resources, and Rebellion in South Asia." *International Security* .

Thyne, Clayton L. 2006. "ABCs, 123's and the Golden Rule: The Pacifying Effect of Education on Civil War, 1890-1999." *International Studies Quarterly* 50:733–754.

Toft, Monica Duffy. 2002. "Indivisible Territory, Geographic Concentration, and Ethnic War." *Security Studies* 12(2):82–119.

Toft, Monica Duffy. 2003. *The Geography of Ethnic Violence: Identity, Interests, and the Indivisibility of Territory*. Princeton, NJ: Princeton University Press.

Walter, Barbara F. 2004. "Does Conflict Beget Conflict? Explaining Recurring Civil War." *Journal of Conflict Resolution* 41(3):371–388.

Weidmann, Nils. 2009. "Geography as Motivation and Opportunity: Group Concentration and Ethnic Conflict." *Journal of Conflict Resolution* 53(4):526–543.

Weinstein, Jeremy M. 2007. *Inside Rebellion.* Cambridge, UK: Cambridge University Press.

Wimmer, Andreas, Lars-Erik Cederman and Brian Min. 2009. "Ethnic Politics and Armed Conflict: A Configurational Analysis of a New Global Dataset." *American Sociological Review* 74:316–337.

Wolitzky, Alexander. 2013. "Cooperation with Network Monitoring." *The Review of Economic Studies* 80(1):395–427.

Wood, Elisabeth. 2003. *Insurgent Collective Action and Civil War in El Salvador.* New York: Cambridge University Press.

Zukerman-Daly, Sarah. 2012. "Organizational Legacies of Violence: Conditions favoring insurgency onset in Colombia, 1964-1984." *Journal of Peace Research* 49(3):473–491.