Self-Managing Terror: Resolving Agency Problems With Diverse Teams

Peter Schram*

January 11, 2021

Abstract

I examine a principal-agents model of subversion with externalities to illustrate a novel mechanism for why diversity can be valuable to organizations: teams of diverse agents can self-manage and discourage their teammates from subversion through compromise. In contrast to standard "ally-principle" type results, I find that integrating more extreme agents can result in better-behaved teams. The model describes, among other cases, radical Islamist terror groups that use foreign fighters. Because foreign and domestic fighters have conflicting preferences over how they want to subvert, integrated teams may self-manage with efficiency gains for the principal. This model explains variation in agency problems and foreign fighter usage in major insurgent groups, including al Qaeda in Iraq, the Haqqani Network, and the Islamic State. Additionally, the theory here can explain management practices in a wide range of alternate settings, for example, where a busy or constrained principal cannot easily implement auditing or incentive contracts.

---

*Assistant Professor, Vanderbilt University, peter.schram@vanderbilt.edu, website This paper has benefited from suggestions and feedback from: Saumitra Jha, James Fearon, Jacob Shapiro, Jonathan Bendor, Nick Eubank, Ken Shotts, Mattias Polborn, Brad Smith, Brenton Kenkel, Brett Benson, Nathan Atkinson, Scott Gates, Steven Callander, Dan Alexander, Constance Wilhelm, two anonymous referees, as well as seminar participants at the Stanford Graduate School of Business, NYU, the Society for Institutional and Organizational Economics Annual Conference, and Vanderbilt University. This paper benefited from outstanding research assistance from Jack Cramer, Ethan Sun, John Konicki, and Haonan Dong. Any opinions, findings, conclusions, recommendations, or errors in the document are those of the author.

In the early 1980s, the Haqqani Network faced an existential challenge. As a major actor in the multi-party insurgency against the Soviet-backed government of Afghanistan, the Haqqani Network needed to conduct disciplined operations against both the government and rival local groups over a vast geographic area while facing intense counterinsurgency pressure. In response, the Haqqani Network became a diverse organization. During this conflict, large numbers of Arab fighters traveled to Afghanistan to fight against the Soviet Union as mujahadeen. While these foreign fighters were viewed cautiously by many Afghan insurgent groups due to their extreme ideology, the Haqqani Network recruited them to create integrated fighting columns of Afghans and foreign fighters (Hamid and Farrall, 2015, 65-167; Brown and Rassler, 2013). And, as an integrated organization, the Haqqani Network has done remarkably well; the Haqqanis have persevered despite nearly four decades of attempts by local actors and global superpowers (the Soviet Union and United States) to destroy the group. In one of the least developed and most conflict prone areas in the world, the Haqqani Network discovered the value of a diverse workforce.

Organizational economics offers explanations for why diversity is valuable to the Haqqani Network. Foreign fighters might introduce new skills or new perspectives on problem solving (Lazear, 1999; Hong and Page, 2001). Alternatively, foreign fighters might provide needed manpower or may be better fighters than local agents. Or, in light of "ally principle" type results in the literature on agency problems (Bendor *et al.*, 2001), the preferences of foreign fighters might be better aligned with leadership than the preferences of domestic agents. If any of these explanations were correct, we would expect similar militant groups to welcome foreign fighters as well. Instead, there is significant variation in foreign fighter use among prominent jihadist groups. The Haqqani Network has integrated foreign fighters for almost 40 years. Al Shaabab, a militant group in Somalia, accepted foreign fighters for several years before turning a blind eye when local fighters began killing off the foreign fighters (Scahill, 2015). Al Qaeda in Iraq (AQI) similarly recruited foreign fighters for several years before deciding in 2007 to start turning them away (CTC, 2007a). Then, when AQI re-emerged as Daesh (or the Islamic State or ISIS) in 2014, the group reversed course again, undertaking the largest recruitment of foreign fighters in history and taking care to integrate foreign fighters into all levels of the organization (Fishman, 2016). This variation in foreign fighter use merits the following questions: why and when is diversity valuable to militant groups?

Diverse preferences among agents are valuable because they present a solution to a critical organizational design problem faced by insurgent leaders: insurgent leadership must design effective teams from imperfect agents to operate in complex environments where it is difficult to implement "top-down" interventions to motivate agents to behave appropriately. The Haqqanis discovered that integrated teams of domestic and foreign agents can, through compromise, self-manage their agency problems more effectively than homogeneous teams of domestic agents, even when foreign agents have preferences that are *less* aligned with the preferences of the leadership relative to domestic agents. Put another way, in contrast to standard ally principle results, the Haqqanis discovered that by adding "worse" agents, strategic interactions between different types of imperfect agents can lead to more efficient teams. This paper analyzes this organizational design problem through a principal-agents model of subversion with externalities between agents.

Jihadist militant groups use teams of imperfect agents. Conducting an insurgency is a multifaceted task; to succeed, the insurgent group must both overthrow the government and emerge as the dominant insurgent group. In aggregate, leadership wants agents to balance attacking both government and rival local actors, but what precisely a team of agents should be doing in a given location at a given point in time is determined by complex and evolving local circumstances (e.g., are rival domestic groups cooperating with or engaging the insurgent group? How active are government forces in an area?). While domestic agents in an insurgent group want the group to succeed, they are also typically motivated by their connection to the local population. That local fighters pursue local power, greed, grievance, or personal security at the expense of the insurgency movement is consistent with extensive empirical evidence and historical anecdotes (Weinstein, 2006; Kalyvas, 2006; Enders and Jindapon, 2010; Shapiro, 2013; Schram, 2019). Foreign fighters are also imperfect, but for different reasons. Jihadist foreign fighters are typically religious extremists who fight to protect the Muslim nation or *umma* when it faces external threats. These preferences make foreign fighters enthusiastic to engage government or non-Islamist forces, but less willing to engage in the local politicking and violence against rival Islamist groups that are necessary to ensure their insurgent group emerges dominant (Hafez, 2010; Hegghammer, 2010; Brown and Rassler, 2013; Schram, 2019). Altogether, local circumstances sometimes make the leadership's preferences more aligned with domestic fighters – when circumstances suggest pursuing a local monopoly on power is most important – while at other times make the principal's preferences more aligned with foreign agents – when circumstances suggest attacking government forces is most important.

The organizational design problem insurgent groups face is particularly hard. When fighting an effective counterinsurgent force, insurgent leadership faces significant risks to conducting oversight. While it is common in principal-agent models to assume that the leadership can observe the outcomes of agents' actions, in an insurgency, if the leadership monitors the agents' operations or attempts to understand the agents' operating environment, then the leadership faces a significant risk of exposing the agents or of the leadership being captured or killed. Altogether, insurgent leadership sends teams of imperfect agents into operating theaters where, ideally, over an extended period of time and with little direct oversight, agents are willing to conduct a wide range of operations in response to varying local circumstances; and, in this setting, unless agents are willing to act against their self-interest, agents will subvert, undermine the organization (Berman, 2009; Shapiro, 2013), and potentially destroy the group (Fishman, 2009).

In this setting, diverse teams can self-manage their inclination to subvert because different kinds of agents have different kinds of misaligned preferences. These different misaligned preferences imply that agents of the same type benefit when their like-minded teammates subvert, but agents of the other type suffer.[1] The agent-agent externality structure used here is different than existing models of shirking (in which agents exert less effort or allocate fewer funds than what the principal would prefer) in terror groups, where

[1] Practically, because pursuing greed or grievance (for local fighters) or engaging government or Western security forces (for foreign fighters) has spillover effects for proximate agents, agents will benefit when their like-minded teammates subvert.

agency problems have negative spillovers on proximate agents (Baccara and Bar-Isaac, 2008; Enders and Jindapon, 2010). That an agent's actions have different effects on different types of proximate agents forms the basis for compromise and is critical to the results below. In a homogeneous team of agents, agents are collectively incentivized to subvert, as each agent benefits from their like-minded teammates' misbehaviors. In contrast, a heterogeneous team includes agents with divergent preferences over the actions that they want to pursue. Because the preferences for subversion on a heterogeneous team pull in different directions, agents may be willing to collectively forgo misbehaving and instead do what is best for the group.

This dynamic can be illustrated in a simple model of agents interacting within homogeneous and heterogeneous teams.[2] Consider an infinite-horizon game in which a two-agent team conducts operations. Let $t \in \{1, 2, 3...\}$ denote periods. In each period, nature selects the state of the world $\omega_t \in \{d, f\}$, then each agent observes ω_t and selects action $x_t \in \{d, f\}$. The state of the world identifies the action that the leadership wants the agents to undertake. Agents have type $\tau \in \{D, F\}$, where agents of type $\tau = D$ ($\tau = F$) prefer action $x_t = d$ ($x_t = f$). Agents receive 2 utils when they perform their preferred action, 2 utils when their teammates perform their preferred action, and 1 util when they perform the action that the leadership wants them to undertake. Assume nature selects $\omega_t = d$ with probability 0.6; this implies the leadership's preferences are more aligned with the preferences of type D agents.

Below I depict the normal forms of the stage game for a homogeneous team of agents of type D. Each normal-form game references the stage game under different states of the world.

Homogeneous Team, $\omega_t = d$

$D \backslash D$	$x_t = d$	$x_t = f$
$x_t = d$	5,5	3,2
$x_t = f$	2,3	0,0

Homogeneous Team, $\omega_t = f$

$D \backslash D$	$x_t = d$	$x_t = f$
$x_t = d$	4,4	2,3
$x_t = f$	3,2	1,1

[2] This simple model was previously introduced in Schram (2019).

In both states of the world, it is a Nash equilibrium for both type D agents to always select the actions that match their type ($x_t = d$, for all $t \in \{0, 1, 2, ...\}$). At times this action is aligned with the preference of the leadership (when $\omega_t = d$), but at other times these agents are subverting (when $\omega_t = f$). Similar results hold for two type F agents always setting $x_t = f$.

However, when a diverse team is formed, a new dynamic can arise. Below is the normal form of the per-period game for a heterogeneous team.

Heterogeneous Team, $\omega_t = d$

D\F	$x_t = d$	$x_t = f$
$x_t = d$	5,1	3,2
$x_t = f$	2,3	0,4

Heterogeneous Team, $\omega_t = f$

D\F	$x_t = d$	$x_t = f$
$x_t = d$	4,0	2,3
$x_t = f$	3,2	1,5

Under both states of the world, it is a Nash equilibrium for agents to select their preferred action (type D chooses $x_t = d$ and type F chooses $x_t = f$) for all t. Under this equilibrium, agents select the actions they prefer, to the detriment of their partner and the group. However, for a sufficiently high discount rate ($\delta \geq 0.8\bar{3}$), using a Nash reversion punishment strategy, an alternate, subgame perfect Pareto improving equilibrium exists where agents select the leadership's most preferred actions ($\omega_t = x_t$ for all t). Essentially, the threat of an agent's teammate subverting for all time can induce better outcomes for the principal and agents. This simple model demonstrates that if agents' preferences for misbehavior are pulling in different directions, they may find it best to coordinate and perform the actions that the leadership would want without the need for oversight by the principal.

This simple model also speaks to when diverse preferences among agents are valuable in the insurgency setting. Comparing the agents' utilities across team structures shows that agents do worse on heterogeneous teams, implying that agents will form homogeneous teams (and subvert) if given the chance. Thus, for agents with diverse preferences to be used effectively (and thus be valuable), the leadership must be able to oversee team formation; if the leadership cannot oversee team formation, the leadership may find it best to exclude the agents whose preferences are least aligned. For insurgent groups, the availability of safe havens approximates the leadership's ability to oversee team formation. Over the course of the group's history, the Haqqani Network possessed a safe haven either in eastern Afghanistan or western Pakistan, which allowed the leaders to organize mixed fighting columns before sending these teams to conduct operations throughout Afghanistan with little direct oversight (Dressler, 2010; Brown and Rassler, 2013). In 2006, AQI's operational bases were repeatedly overrun, forcing the group to decentralize command authority. As a result, AQI's domestic fighters chose not to work with the foreign fighters, instead leaving foreign fighters behind in safe-houses while local fighters stole from, killed, and generally alienated the Iraqi population. Eventually, the leadership began turning foreign fighters away (CTC, 2007a; Fishman, 2009). Several years later, when AQI became the Islamic State (ISIS) and successfully occupied territory in Iraq and Syria, the group began soliciting foreign fighters and forming integrated operational units again (Weiss, 2015; Gates and Podder, 2015).

While I analyze self-managing teams through the lens of jihadist militant groups, the theory is not limited to this specific case. The theory suggests that when different agents have different preferences for subversion, diverse teams can self-regulate with efficiency gains for the principal. The intuition for this organizational solution to agency problems is fairly natural; if an agent has a penchant for one type of misbehavior—say corruption—an effective leader would not put that agent on a team with other corrupt agents. While this is far from a model for all principal-agent situations, outside of the insurgency setting, this theory can help explain the value of diversity in corporate boards (Adams *et al.*, 2010), government agencies like the Federal Trade Commission (Wilson, 1989), and police forces (Maciag, 2015).

I highlight three primary contributions. First, I present a novel explanation for how diversity can benefit organizations. Lazear (1999) suggests that the existence of the "global firm" is in itself a puzzle, as bringing together a multinational workforce under a single organization imposes a range of cultural costs and tensions. This paper suggests that, in some settings, these tensions can be what makes diversity valuable. That teams of diverse agents can self-regulate their agency problems is distinct from existing hypotheses for how diversity can facilitate productivity or team problem solving (Lazear, 1999; Hong and Page, 2001). And while other results illustrate that diversity among agents can result in better outcomes for the organization, existing results are borne from legislative signaling settings with active principals (Battaglini, 2002; Hirsch and Shotts, 2015), from direct competition between agents (Bonatti and Rantakari, 2016), or from environments where diversity improves organizational productivity through reducing free-riding problems (Baccara and Yariv, 2016). The result here is that diversity can be a simple, "hands-off" tool for addressing moral hazard by encouraging compromise, which is more consistent with cases (like the insurgency setting) where agents have limited interactions with the principal, are not competing, and dislike working with members of the outgroup. Second, within the context of self-managing teams, I assess the natural intuition that the principal wants agents whose preferences are more aligned with the preferences of the principal. I find this intuition does not hold. Rather, teams self-manage best when agents' preferences "offset" one another, meaning the principal may recruit an agent whose preferences are less like the principal's to counterbalance the preferences of that agent's teammate. Third, I show that creating diverse teams functions better in low-information and bureaucratic settings than other common contracting techniques.

1 Motivating the Model

1.1 Insurgent Preferences and Importance of the Topic

This discussion provides some brief background to justify the utility functions in the general model and describes the importance of the foreign fighter phenomenon. Within jihadist militant groups participating in multi-party insurgencies, there are three distinct groups of actors that each possess distinct preferences: the leadership, foreign agents, and domestic agents.

To be successful, an insurgent group must both overthrow the government and, should the government fall, be positioned to consolidate political power; to accomplish these goals, the group must balance attacking government forces and rival civilian or rival insurgent forces (Whiteside, 2016). However, what precisely agents should be doing in a given location is determined by local circumstances that agents (who are operating at the location) would observe, but the leadership (who are isolated from day-to-day operations) would not observe. For example, if rival insurgent groups were attacking Haqqani Network agents, the Haqqani Network's leadership would want its agents to respond by attacking these rivals. However, even if the leadership observes its agents attacking local actors, the leadership does not easily know if the agents are being attacked and are rightly responding, or if the agents are pursuing local power and alienating civilians at the expense of the group's ultimate goal.

Domestic fighters have goals similar to that of the leadership. However, they tend to care more about gaining a local monopoly on power in the short term than leadership does. I will use the term "domestic fighters" to classify those fighters that possess a pre-existing social network and connection to the local population.[3] Through such common insurgent activities as the enforcement of Sharia law, smuggling, and racketeering (Moghadam and Fishman, 2010; Shapiro, 2013), domestic agents can settle old grievances, protect their social network, and pursue wealth to an extent that ideologically driven outsiders (foreign fighters) or the group's leadership both could and would not. That local fighters pursue greed, grievance, or personal security at the expense of the insurgency movement, is consistent with anecdotal evidence from both AQI and the Haqqani Network, as well as with existing literature on agency problems within domestic insurgencies (Weinstein, 2006; Kalyvas, 2006; Hamid and Farrall, 2015; CTC, 2007a). Additionally, this perspective is not limited to radical Islamist groups. Trotsky (1971) claimed "local cretinism is history's curse on all

[3] This classification is better than considering, for example, nationality, as actors from North Waziristan in Pakistan will have more in common with civilians in Eastern Afghanistan (where the Haqqani Network operates) relative to Afghans from the West or North.

peasant riots" and Mao (1938) criticized the peasant guerrilla units "which are frequently preoccupied with local considerations to the neglect of the general interest."

Islamist foreign fighters, meanwhile, are ideologues who travel distances to conflict zones like Iraq in 2005 or Afghanistan in 2001 not because they have a personal connection to the conflict (for example, they are not just co-ethnic) but rather because they believe it is their religious duty to protect the Muslim nation (the *umma*) when it faces external threats such as the one posed by Western forces or Western-backed governments (Hegghammer, 2013). While foreign fighters also undoubtedly benefit from their insurgent group's success, foreign fighters prefer to engage Western security forces or non-radical Islamist governments (e.g., the Afghan government under Hamid Karzai or the Iraqi government under Nouri al-Maliki) than to engage co-religious militants or civilians and to become involved in local politicking (Hafez, 2010). Secondary documents on foreign fighter ideology and recruitment patterns (Felter and Fishman, 2007; Hafez, 2010; Kirdar, 2011) and internal documents discussing the motivations and religious devotion of foreign fighters (CTC, 2007a) all support this view.

Thus, leadership possesses preferences that are sometimes more in line with the preferences of foreign fighters and sometimes more in line with those of local fighters, depending on the specific, complex, and evolving local circumstances that the leadership cannot easily discern. Overall, I assume leadership preferences are closer to those of domestic fighters because both groups share a desire to consolidate their power after competing parties are defeated. In contrast, foreign fighters are generally less interested in securing the group's long run success and care more that Western or apostate government forces are defeated (Hegghammer, 2010; McChrystal, 2013). Should the militant group be successful, foreign fighters often move on to the next battle zone. This assumption has empirical significance: because foreign fighters are less aligned with the preferences of the leadership, if forced to choose, leadership will work with local fighters rather than foreign fighters (as happened in AQI and al Shabaab).

The empirical phenomena at the center of this paper — radical Islamist foreign fighters being brought into insurgent groups—is a defining characteristic of contemporary civil wars (Walter, 2017). Of the 101 conflicts listed in the 2010-2019 period within the UCDP/PRIO dataset (Pettersson and Öberg, 2020), 31 of these conflicts had radical Islamist foreign fighters fighting alongside radical domestic insurgents.[4] This figure is not the full story, as foreign fighters are more common in violent insurgencies; of the ten most deadly conflicts occurring within this period, nine of them had foreign fighters. Additionally, over this period, the UCDP/PRIO dataset reports that conflicts without foreign fighters generated 69,204 battle deaths, while conflicts with radical Islamist foreign fighters generated 597,665 conflict deaths. While it is impossible to identify the true impact of foreign fighters in these conflicts, these figures suggest that if we have an interest in understanding the particulars of the most violent contemporary insurgencies, we need to have a better understanding of the role of radical Islamist foreign fighters. Of course, not all foreign fighters are radical Islamists. Foreign volunteers fought on both sides of the Spanish Civil War (1936-1939), and the FARC in Colombia had volunteers from other Latin American countries and Europe.

1.2 Other Applications of Self-Managing Teams

While the emphasis of this paper is radical jihadist groups, the self-managing teams described here apply to cases outside of radical Islamist insurgencies. The intuition for the results is fairly natural: if an insurgent, soldier, or law enforcement agent has a penchant for corruption, then an effective leader would not put that agent on a team with other corrupt agents or else the team may undertake corrupt actions. Alternatively, if a CEO of a global company organized a committee to identify the ideal location for a new factory, then this CEO probably would not make every member of the committee Brazilian out of fear that the committee would be biased towards putting the factory in Brazil. In settings where agents may have many different and conflicting reasons for wanting to subvert, creating diverse teams can foster cooperation with efficiency gains for the principal.

Within the context of security studies, the self-managing teams described here can function without Islamist foreign fighters. Consider insurgencies with strong ideological

[4] I discuss the methodology for these statistics in the Appendix.

components, like the FARC in Colombia, the CPP in the Philippines, and the IRA in Northern Ireland. Within these groups, there will undoubtedly be some members who are more ideologically driven (similar to foreign fighters discussed above) while others are more concerned with local matters and practicalities (similar to domestic fighters). If leadership can create integrated teams of agents with diverse preferences, this paper suggests the group can achieve efficiency gains. Additionally, insurgent (or government) practices like relocating soldiers to operate in conflict theaters that they are less familiar with can introduce a new set of preferences that can offset the preferences of the agents that they are embedded with.[5] Furthermore, third-party interventions are fairly common in civil wars, and could be viewed through the lens of this research. One key factor for why third-parties intervene in civil wars is so that they can interact with and influence the behavior of their ally. And finally, even state militaries that have a general ability to monitor and audit their agents to prevent subversion will, at times, have teams of agents operating without much oversight (for example, in special forces operations) or will not have an ability to issue top-down management. In these scenarios, having a balanced team can lead to better outcomes.

This model can apply to a range of other settings. In the most general sense, effective teams are often described as possessing a range of perspectives and have agents regularly cooperating and engaging in self-sacrifice for the good of the team (Beyerlein and Johnson, 1994; Yeatts and Hyten, 1998). This model supports these observations. More specifically, this model could describe a range of settings, like corporate board composition (see Bathala and Rao (1995) and Adams *et al.* (2010) for examples). Corporate boards set the general strategy of their companies, and are often composed of a mix of outsiders and insiders. While insider-members may have preferences for empire building or preferences against the dissolution of divisions and layoffs, outsider-members who lack strong connections to the employees can offset the preferences of insider-members and potentially coordinate on better firm strategy. Furthermore, this model could describe the behavior of government agencies staffed by multiple types of agents (for example, the economists and lawyers in the

[5]For example, if Insurgent A from Location A is relocated to Location B, then Insurgent A attacking a rival insurgent group in Location B could generate more conflict in Location A, which could hurt Insurgent A's social network.

Federal Trade Commission; see Wilson (1989)). Finally, this model offers a formal rationale for why diversity is seen as valuable to community policing (Ramsey and Robinson, 2015).[6]

The results here apply to subversion settings where constrained leadership must design effective teams from imperfect agents to operate in complex environments. My model best describes cases where leaders face external, bureaucratic, or organizational constraints—such as counterinsurgency pressure, explicit rules on how the leadership can interact with agents, or a massive organization—where the primary interaction for agents is with other teammates rather than with the principal. Naturally, there are important limitations for where the theory can apply. First, this is a model for subversion, not shirking. If one agent's misbehavior does not benefit a teammate (as is the case in Holmstrom (1982)), then this theory does not apply. Second, this is not a model for settings where a principal uses diverse agents to motivate performance through competition, like is common in the legislative signaling setting (Hirsch and Shotts, 2015).

1.3 Related Literature

Since Chai (1993) pioneered an organizational economics approach to terror groups, a growing literature discusses how terror and insurgent groups mitigate their agency problems (Gates, 2002; Weinstein, 2006; Shapiro and Siegel, 2007; Baccara and Bar-Isaac, 2008; Berman and Laitin, 2008; Enders and Jindapon, 2010; Shapiro, 2013). This scholarship explains many puzzling facets of insurgent organization (see Shapiro (2019) for a review), but the analyses almost always focus on a single type of misbehaving agent without considering agent-agent interactions.[7] As such, these papers may be missing relevant interactions that exist in large and diverse militant groups, such as ISIS and the Haqqani Network. This paper not only finds that exploiting agent-agent interactions can be used to resolve agency problems, but also highlights the feasibility of this technique in challenging contracting environments. This paper is similar to Schram (2019), which describes the

[6]One of the terms of the Good Friday Agreement which ended "The Troubles" in Northern Ireland called for fair representation of Catholics within police forces; one way this was implemented was through a 50:50 recruitment policy for Catholics and Protestants from 2001 to 2011 (Caparini and Hwang, 2019).

[7]Baccara and Bar-Isaac (2008) and Enders and Jindapon (2010) are exceptions.

preferences of actors in an insurgency, introduces the simple model that was presented earlier, and describes how counterinsurgency undermined self-managing teams in AQI. However, this paper differentiates itself from Schram with a generalized principal-agents model, a discussion on comparative statics and extensions, an analysis of the principal's problem, and a discussion of non-insurgency applications.

The finding that diversity can be exploited for efficiency gains for the principal has been shown in legislative signaling models (Gilligan and Krehbiel, 1989; Dewatripont and Tirole, 1999; Battaglini, 2002; Hirsch and Shotts, 2015). This paper differentiates itself in two ways. First, this paper examines subversion in a delegation setting rather than a signaling setting. Second, in the signaling models above, after the principal organizes a diverse team of agents, the principal plays a critical role in realizing the efficiency gains. In Gilligan and Krehbiel (1989),Dewatripont and Tirole (1999), and Hirsch and Shotts (2015), the principal is able to assess the quality of a given policy, and in Battaglini (2002), the principal interprets a multidimensional message to construct an optimal policy. Here, in contrast, after the principal organizes a diverse team, the principal relies on agents to do much of the management.

Research on delegation outside of legislative signaling similarly finds violations of the ally principle. Bertelli and Feldmann (2007) and Gailmard and Hammond (2011) find that non-allied agents can better represent the interests of the principal in political bargaining because they can stake out stronger policy positions relative to the agents who share the principal's preferences, or that non-allies can effectively tie the hands of future policymakers. Outside of game structures where multiple agents interact, Bubb and Warren (2014) find that employing biased agents can incentivize the principal to use a review process more, with efficiency gains for the principal. Also, Bendor and Meirowitz (2004) consider a series of models and extensions, showing sometimes delegating to a less aligned agent can produce better information revelation or avoid collective action problems within the principal-agent relationship. The results in this paper are distinct for several reasons. First, while the institutional structures captured in Bertelli and Feldmann, Gailmard and Hammond, and

Bubb and Warren are appropriate for the cases that they analyze and help drive the value of non-allies, this paper shows that non-allies can be valuable in much less structured settings. Second, what makes less-allied agents valuable here is that they can counter the misaligned preferences of their teammates, make subversion hurt not just the principal but also their teammates, and thus incentivize compromise between agents in ways that are productive for the principal.

Finally, beginning with Holmstrom (1982), a series of papers have also examined principal-agents problems with teams. As examples, these papers address topics ranging from how dividing tasks (Holmstrom and Milgrom, 1991), side contracting (Tirole, 1986; Holmstrom and Milgrom, 1990; Che and Yoo, 2001; Jackson and Wilkie, 2005), and externalities (Segal, 1999) impact the principal's problem.[8] This paper also speaks to the politics of organizational decision making, where organizational or institutional factors play a significant role in determining how teams behave (Gibbons and Roberts, 2012). For example, Bonatti and Rantakari (2016) describes how, when agents face costs for producing policies, diverse policy preferences among agents can lead to greater allocations of effort into producing policies.

2 Model

This is a principal-agents model of subversion with externalities between agents. The model has two stages: a first stage where the principal designs the organization, and a second stage where agents repeatedly conduct operations.

In the first stage, the principal defines utility transfers to agents and oversees the formation of a team of agents. The principal has two choices on how to structure transfers: the principal can offer an incentive contract where the amount of the transfer is based on agents' actions, or the principal can simply offer a fixed transfer. In order to condition

[8] I will not attempt to list all works discussing principal-agent modeling of teamwork. See Gibbons and Roberts (2012) for a review.

transfers on the agent's actions, the principal sets $m = 1$ and incurs a "monitoring" cost $\kappa_m > 0$. If the principal sets $m = 0$, the principal can still offer agents a flat utility transfer and does not incur a monitoring cost. The principal then forms a team of agents. Agent 1 is assumed to be a domestic type.[9] The principal sets $o_p \in \{d, f, u\}$ to designate agent 2 as a domestic ($o_p = d$) or foreign type ($o_p = f$), or to remain uncommitted and delegate this decision to agent 1 ($o_p = u$). If the principal sets $o_p = f$ or $o_p = d$, the principal pays a one-time "organization" cost $\kappa_o > 0$. If the principal sets $o_p = u$, then agent 1 selects $o_1 \in \{d, f\}$ to partner with either a domestic type or foreign type agent 2. When transfers and team composition are set, agents then have the option to accept participating in operations ($b_i = a$) or reject participating ($b_i = r$). If either agent selects r, then the game terminates and all actors receive their reservation utilities denoted by R_p and R_a for the principal and agents (respectively).

The second stage is an infinite horizon game where agents repeatedly conduct operations. Time is discrete and indexed by $t \in \{1, 2, 3, ...\}$. At the start of each period t, nature draws a realization of $\omega_t \in [-1, 1]$ which represents what actions the principal wants each of the agents to perform. Each ω_t is drawn independently from a continuous distribution function F with full support where $\mathbb{E}(\omega_t) = 0$. The distribution F is common knowledge and the agents observe ω_t, but the principal does not. After ω_t is realized, both agents simultaneously select actions $a_{i,t} \in \mathbb{R}$ with $i \in \{1, 2\}$. The convexity of the action space captures that in a given period, agents allocate their time to some mixture of activities and that the principal has a most preferred mix of activities (represented by ω_t).[10] In the insurgency setting, agents can spend more time attacking local actors or rival insurgent groups (represented by more negative values of $a_{i,t}$), spend more time attacking government actors or Western forces supporting government actors (represented by more positive values of $a_{i,t}$), or spend time mixing between the two (values of $a_{i,t,}$ close to 0). After agents conduct actions, per-period utilities are realized, and the game moves to period $t = t + 1$.

[9] This is a simplifying assumption, as the principal weakly prefers that agent 1 is the type of agent whose preferences are more in line with the leadership.

[10] An alternate interpretation of the convex action set is that it represents the intensity of attacks on domestic actors or government forces. For example, overtaxing a civilian could be represented by $a_{i,t} = -0.1$, while inflicting excessive harm on a civilian could be $a_{i,t} = -0.9$.

Based on the preferences described in Section 1.1, domestic agents prefer taking actions against local actors or rival insurgent groups, while foreign agents prefer taking actions against Western or government forces. Formally, domestic agents have ideal point χ_d where $\chi_d \leq -1$ and foreign agents have ideal point χ_f where $\chi_f \geq 1$. I assume in expectation that the leadership's preferences are weakly more in line with domestic agents, or $|\chi_d| \leq |\chi_f|$.

For agent $i \in \{1,2\}$ that is type $\tau \in \{d, f\}$, and letting $j \in \{1,2\}$ with $i \neq j$, summed across periods, agent i has utility function

$$U_i = \sum_{t=1}^{\infty} \delta^{t-1} \left(-|a_{i,t} - \chi_\tau| - |a_{j,t} - \chi_\tau| - \gamma|a_{i,t} - \omega_t| + G_{i,t}\right). \tag{1}$$

I let $\delta \in (0,1)$ denote the common discount factor and $\gamma > 0$ is a constant. I assume agents incur disutility when they and their partners select actions that deviate from their ideal points, as represented in the linear[11] loss terms $-|a_{i,t} - \chi_\tau|$ and $-|a_{j,t} - \chi_\tau|$.[12] Practically, these terms imply that domestic agents incur disutility when they and their teammates are not pursuing local power, and foreign agents incur disutility when they and their teammates are not engaging Western forces. I also assume that when agents deviate from what the principal would want them to do, they incur disutility, as represented in the $-\gamma|a_{i,t} - \omega_t|$ term. Practically, when agents subvert, they are inappropriately attacking actors, fostering new hostilities, or generally undertaking actions that have negative ramifications for the group, which would have a negative impact on the misbehaving agents.[13] I assume $1 > \gamma$, which implies agents are motivated to subvert. The $G_{i,t}$ function denotes the per-period utility transfer from the principal to agent i. I limit the analysis to contracting schedules that, for a transfer in period t, do not rely on events or information outside of what occurred

[11] I assume linear utility functions to allow the principal to achieve the first-best outcome when using incentive contracts. Linear utilities here "stack the deck" in favor of using incentive contracts, which makes them a more competitive benchmark to creating self-managing teams.

[12] In the online appendix I include a more nuanced utility function, where each agent has utility $U_i = \sum_{t=1}^{\infty} \delta^{t-1} \left(-\alpha|a_{i,t} - \chi_\tau| - \beta|a_{j,t} - \chi_\tau| - \gamma|a_{i,t} - \omega_t| + G_{i,t}\right)$, where α and β are positive constants and $\alpha > \gamma$.

[13] It might also be expected that one agent's subversion would hurt that agent's teammate. This assumption would make heterogeneous teams more willing to self-manage their agency problems.

in period t. The transfer function can then be defined as mapping $G_{i,t} : \{\emptyset\} \cup \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$.

The principal has utility function

$$U_p = \sum_{t=1}^{\infty} \delta^{t-1} \left(-|a_{1,t} - \omega_t| - |a_{2,t} - \omega_t| - G_{1,t} - G_{2,t} - m\kappa_m\right) - 1_{o\in\{d,f\}}\kappa_o. \tag{2}$$

Regarding contracting, the principal may choose to pay incentive contract and oversight costs, as represented in the $G_{1,t}$, $G_{2,t}$, $m\kappa_m$, and $1_{o\in\{d,f\}}\kappa_o$ terms. Regarding the behavior of the team, the principal prefers that both agents set their actions $a_{i,t} = a_{j,t} = \omega_t$. As is modeled, each agent faces the state of the world independently, and individual deviations from the state hurt both the principal and agents. This is different from, for example, both agents conducting actions that are averaged or aggregated and then compared to the state of the world. I examine the non-averaged model here because two different instances of subversion (i.e. two wrongs) should not necessarily be treated as a team not subverting (i.e. making a right).[14] I discuss what the model would look like with averaged actions in the extensions below, and I show that cooperation among diverse agents can still lead to efficiency gains for the principal. Also, it is worthwhile to mention that the model explored here is substantively equivalent to both agents on a team facing their own state of the world variable.

Assumption 0 summarizes the above assumptions on preferences.

Assumption 0: $\gamma \in (0, 1)$. Realizations of the principal's preferred activities are represented by $\omega_t \in [-1, 1]$, where $\mathbb{E}(\omega_t) = 0$. The agent's ideal points correspond to these points by $\chi_d \leq -1 < \mathbb{E}(\omega_t) = 0 < 1 \leq \chi_f$, with $|\chi_d| \leq |\chi_f|$.

The game can be summarized as follows.

1. The principal selects monitoring decision $m \in 0, 1$, organization decision $o_p \in \{d, f, u\}$,

[14] For example, if $\omega_t = 0$, I am not assuming that the principal is indifferent between action pair $a_{i,t} = -1$ and $a_{j,t} = 1$ and action pair $a_{i,t} = a_{j,t} = 0$.

and transfers $G_{1,t} : \{\emptyset\} \cup \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ and $G_{2,t} : \{\emptyset\} \cup \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ for all t.

2. If the principal set $o_P \in \{d, f\}$, the game proceeds to 3. If the principal set $o_P = u$, agent 1 selects if agent 2 is domestic or foreign $o_a \in \{d, f\}$.

3. Having observed the contracts, each agent accepts being in the group and sets $b_i = a$ or rejects the group and sets $b_i = r$. If either agent rejects the group, then the game terminates and all actors receive their reservation values (R_p and R_a).

4. Period $t = 1$ begins.

5. Nature designates $\omega_t \in [-1, 1]$, which is observed by the agents.

6. Agents 1 and 2 simultaneously select actions $a_{1,t} \in \mathbb{R}$ and $a_{2,t} \in \mathbb{R}$.

7. Utilities are realized, the game returns to 5, and t updates to $t + 1$.

I limit my analysis to subgame perfect equilibria. Even so, multiple equilibria can exist in the repeated second stage. To further limit the set of subgame perfect equilibria, I introduce three criteria for equilibrium selection. First, I will only consider equilibria supported by Nash reversion. Second, I will only consider a type of subgame perfect equilibrium that I refer to as a "shading equilibrium."

Definition: A subgame perfect equilibrium is a "**shading equilibrium**" if, for $z_i \in [0, 1]$, agent i of type τ selects actions $a_{i,t} = z_i\omega_t + (1 - z_i)\chi_\tau$ for all t.

Each agent's per-period actions follow from the above affine function of the state variable, their ideal point, and the choice variable z_i, and are restricted to fall between the agent's ideal point and the state variable. In other words, in a shading equilibrium, agents "shade" some fixed proportion z_i towards the principal's ideal point at ω_t from their ideal point at χ_d or χ_f. Subgame perfect equilibria that are not shading equilibria include equilibria when agents vary their behavior across periods[15] or select from different

[15] For example, agents select $z_i = 0.6$ on even periods and $z_i = 0.4$ on odd periods.

functional forms based on the state variable.[16] That agents only select values between their ideal points (χ_d and ω_t or χ_f and ω_t) is in place to simplify analysis. I relax this assumption in the "Other Extensions" section, and the results do not substantively change.

Regardless of whether a domestic-domestic or domestic-foreign team is formed, when the principal is not using utility transfers, both agents setting $z_1 = z_2 = 0$ always exists as a shading equilibrium. However, this may not be the only shading equilibrium. The third criterion resolves this. I also assume that agents select the shading equilibria in which agents "shade" the most towards the principal's ideal point ω_t. All three criterion are summarized in Assumption 1.

Assumption 1: Agents will select the shading equilibrium that is supported by Nash reversion and that is characterized by (z_1, z_2) where
$(z_1, z_2) \in \arg\max_{(z_1,z_2)\in Z^*} \{-|a_{1,t}(z_1, \omega_t) - \omega_t| - |a_{2,t}(z_2, \omega_t) - \omega_t|\}$.

Being cognizant of Folk Theorem results in repeated games, readers may be worried that Assumption 1 induces agents to select an odd equilibrium in which the equilibrium's peculiarities are necessary for heterogeneous self-managing teams to function. This is not the case. Limiting the analysis to shading equilibria imposes a simple structure to equilibria analysis. Limiting analysis to equilibria supported by Nash reversion eliminates potentially implausible equilibria that rely on extreme off-path punishments. This criterion also means that any selected equilibrium will be weakly Pareto improving for the agents relative to the $z_1 = z_2 = 0$ shading equilibrium (where agents match their actions to their ideal points). And, while limiting analysis to equilibria where agents select the actions that are closest to the principal's ideal point may seem strong, this is akin to assuming that the principal can "nudge" agents who are deciding between multiple equilibria into the one that is good for the organization (and that is weakly better for the agents from the $z_1 = z_2 = 0$ shading equilibrium). This seems like a valid assumption: by virtue of being the leader of a large,

[16] For example, when $\omega_t \leq 0$ domestic agents select $a_{i,t} = \omega_t - 0.5(\omega_t - \chi_d)$ and when $\omega_t > 0$ domestic agents select $a_{i,t} = 0.1 + \omega_t - 0.5(\omega_t - \chi_d)$.

successful militant group, leadership probably has some managerial ability to convince agents not to play destructive equilibria. Also, in the Other Extensions section, I relax Assumption 1 by considering alternatives to shading equilibria (as defined here) and the case where agents select actions that maximize the team's joint utility;[17] these changes do not substantively change the results.

3 How Diverse Teams of Agents Behave

Before discussing the techniques the principal can use, I first describe the mechanics of how a diverse team will self-manage in the repeated second stage. On a diverse team, a shading equilibrium always exists where agents match their actions to their respective ideal points, or $a_{1,t} = \chi_d$ and $a_{2,t} = \chi_f$ for all t (where $z_1 = z_2 = 0$). To arrive at the equilibrium that satisfies the maximization condition in Assumption 1, the $z_1 = z_2 = 0$ equilibrium acts as the Nash reversion punishment-phase that players would enter upon observing deviations from equilibrium behavior. Following Assumption 1, within a heterogeneous team, agents select actions $a_{1,t} = (1 - \tilde{z}_1)\chi_d + \tilde{z}_1\omega_t$ and $a_{2,t} = \tilde{z}_2\omega_t + (1 - \tilde{z}_2)\chi_f$ for all t, with $\tilde{z}_1$ and $\tilde{z}_2$ defined below.

Definition: $\tilde{z}_1$ and $\tilde{z}_2$ are defined as as

- $\tilde{z}_1 = 1$ and $\tilde{z}_2 = 1$ if $\tilde{k}_f \geq 1$,
- $\tilde{z}_1 = 1$ and $\tilde{z}_2 = \tilde{k}_f$ if $\tilde{k}_d\tilde{k}_f \geq 1$ and $\tilde{k}_f < 1$, and
- $\tilde{z}_1 = 0$ and $\tilde{z}_2 = 0$ if $\tilde{k}_d\tilde{k}_f < 1$,

where $\tilde{k}_d = \frac{\delta\chi_f}{(1-\gamma)(1-\delta-\chi_d)}$ and $\tilde{k}_f = \frac{-\delta\chi_d}{(1-\gamma)(\chi_f+1-\delta)}$.

I derive $\tilde{z}_1$ and $\tilde{z}_2$ in the Appendix. To provide intuition, I plot $\tilde{z}_1$ and $\tilde{z}_2$ relative to χ_d in Figure 1. Note that the expressions $\tilde{k}_f$ and $\tilde{k}_d\tilde{k}_f$ are both decreasing in χ_d.

[17] There is no way to identify the single "best" equilibrium for the agents because, on a heterogeneous team, there can exist multiple Pareto efficient equilibria.

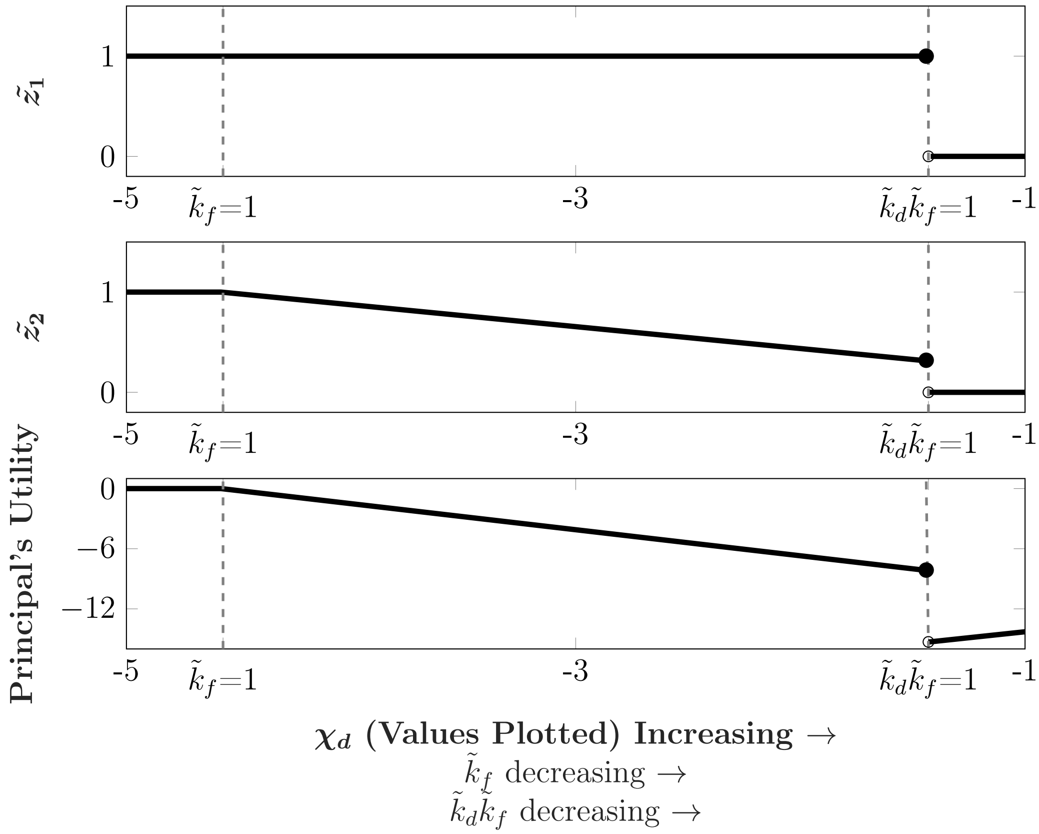

Figure 1: Plots of $\tilde{z}_1$, $\tilde{z}_2$, and the principal's utility against values of χ_d.

Notes: Parameter values are $\gamma = 0.51$, $\delta = 0.58$, $\chi_f = 5$, and $\chi_d \in [-5, -1]$. The "Principal's Utility" refers to the principal's expected utility from the agents' actions. The expressions $\tilde{k}_f$ and $\tilde{k}_d\tilde{k}_f$ are both decreasing in χ_d, which implies that: (a) in the region left of the $\tilde{k}_f = 1$ border, $\tilde{k}_d > 1$ and $\tilde{k}_f > 1$ holds; (b) in the region right of the $\tilde{k}_d\tilde{k}_f = 1$ border, $\tilde{k}_d\tilde{k}_f < 1$ holds; and (c) in the region in between, $\tilde{k}_f \leq 1$ and $\tilde{k}_d\tilde{k}_f \geq 1$ holds.

Agents are willing to shade up to levels $z_1 \leq min\left\{1, z_2\tilde{k}_d\right\}$ and $z_2 \leq min\left\{1, z_1\tilde{k}_f\right\}$. Due to the maximization condition in Assumption 1, these inequalities will hold with equality. That each agent's willingness to shade is an increasing function of their teammates shading level creates three general possibilities for the final selected shading level; to illustrate why, it is useful to compare the case when $\tilde{k}_d < 1$ and $\tilde{k}_f < 1$ to the case when $\tilde{k}_f \geq 1$ (which, by Assumption 0, implies that $\tilde{k}_d \geq 0$). When $\tilde{k}_f \geq 1$ (the portion of Figure 1 to the left of $\tilde{k}_f = 1$), each agent is willing to shade at a level weakly greater than that of their teammates, resulting in $\tilde{z}_1 = \tilde{z}_2 = 1$ as the selected equilibrium shading levels. In contrast, when $\tilde{k}_d < 1$ and $\tilde{k}_f < 1$ (which occurs for the smallest values of D's ideal point χ_d within portion of Figure 1 to the right of $\tilde{k}_d\tilde{k}_f = 1$), each agent is only willing to shade a fraction (i.e. smaller than one) of their teammate's selected level of shading, making $\tilde{z}_1 = \tilde{z}_2 = 0$ the only possible shading equilibrium. The equilibrium behavior between these parameter spaces is dictated by whether $\tilde{k}_d\tilde{k}_f \geq 1$ or $\tilde{k}_d\tilde{k}_f < 1$, which is the cut point where non-zero shading levels can (when $\tilde{k}_d\tilde{k}_f \geq 1$) or cannot (when $\tilde{k}_d\tilde{k}_f < 1$) be supported. Thus, referencing the bullet points: when $\tilde{k}_f \geq 1$, agents set $\tilde{z}_1 = 1$ and $\tilde{z}_2 = 1$, meaning they are completely self-managing their agency problems; when $\tilde{k}_d\tilde{k}_f \geq 1$ and $\tilde{k}_f < 1$, agents set $\tilde{z}_1 = 1$ and $\tilde{z}_2 = \tilde{k}_f$, meaning agent 1 matches their action to the principal's most preferred action, but agent 2 only partially self-manages;[18] when $\tilde{k}_d\tilde{k}_f < 1$ then no non-zero level of shading can be supported.

Based on $\tilde{z}_1$ and $\tilde{z}_2$, I can discuss comparative statics on the principal's expected utility. The most important comparative statics to consider are those when agents are actually self-managing, which occurs when $\tilde{k}_d\tilde{k}_f \geq 1$, or when changes in parameters induces agents to shift from not self-managing to self-managing, which occurs when $\tilde{k}_d\tilde{k}_f < 1$ shifts to $\tilde{k}_d\tilde{k}_f \geq 1$.

Observation 1: Within a heterogeneous team:

- within the region where $\tilde{k}_d\tilde{k}_f \geq 1$, the principal's expected utility is weakly increasing in γ, and weakly decreasing in χ_d and χ_f.
- the expression $\tilde{k}_d\tilde{k}_f$ is increasing γ, and decreasing in χ_d. If a change in γ, or χ_d induces a change from $\tilde{k}_d\tilde{k}_f < 1$ to $\tilde{k}_d\tilde{k}_f \geq 1$ (or from $\tilde{k}_d\tilde{k}_f \geq 1$ to $\tilde{k}_d\tilde{k}_f < 1$), then the principal's expected utility is strictly increasing (or strictly decreasing) in that variable.
- the expression $\tilde{k}_d\tilde{k}_f$ is increasing in χ_f. If a change in χ_f induces a change from $\tilde{k}_d\tilde{k}_f < 1$ to $\tilde{k}_d\tilde{k}_f \geq 1$ or from $\tilde{k}_d\tilde{k}_f \geq 1$ to $\tilde{k}_d\tilde{k}_f < 1$, the effects on the principal's utility are ambiguous.
- within the region where $\tilde{k}_d\tilde{k}_f < 1$, the principal's expected utility is unchanging in γ,

[18]Because $|\chi_d| \leq |\chi_f|$, agent 1 is always willing to select a level of shading $\tilde{z}_1$ that is (weakly) greater than $\tilde{z}_2$.

strictly increasing in χ_d, and strictly decreasing in χ_f.

The most surprising result is that, when $\tilde{k}_d\tilde{k}_f \geq 1$, the principal's expected utility is weakly decreasing in χ_d. As shown in Figure 1, when χ_d decreases—when the domestic agent's ideal point is further from the set of actions that the principal wants the agent to conduct—the team will weakly shade more towards the principal's ideal actions, with weak utility gains for the principal. This result contrasts with standard ally principle results, which suggest that principals prefer agents whose ideal points are closer to their own preferences over agents whose ideal points are further from their own preferences. Rather, this result shows that the closer agent 1's ideal point is to the action the principal wants the agent to conduct, the weakly worse the principal does.

The intuition for why decreasing χ_d can be better for the principal is as follows. Consider what a decrease in χ_d does when $\tilde{k}_d\tilde{k}_f \geq 1$ and $\tilde{k}_f < 1$. By decreasing χ_d, it makes the Nash reversion punishment of $\tilde{z}_1 = \tilde{z}_2 = 0$ worse for the foreign agent (agent 2) because χ_d becomes further from the foreign agent's ideal point χ_f. By making deviations from equilibrium behavior worse, agent 2 is willing to remain in a broader set of non-zero shading equilibria, which is reflected in the increase in $\tilde{k}_f$. Due to the maximization condition in Assumption 1, the increase in $\tilde{k}_f$ is reflected in equilibrium behavior where $\tilde{z}_2 = min\left\{\tilde{k}_f, 1\right\}$. Of course, decreasing χ_d also affects agent 1. As a first order effect, decreasing χ_d means that agent 1 does worse when setting $\tilde{z}_1 = 1$. However, as a second order effect, decreasing χ_d increases $\tilde{z}_2$, which makes remaining on the equilibrium path better for agent 1. In aggregate, decreases in χ_d help support non-zero shading equilibria; taking first order conditions of the $\tilde{k}_d\tilde{k}_f$ expression shows that decreases in χ_d increase $\tilde{k}_d\tilde{k}_f$, meaning that a decrease in χ_d will never break the $\tilde{k}_d\tilde{k}_f \geq 1$ condition, implying that agent 1 is willing to remain at $\tilde{z}_1 = 1$. Altogether, when $\tilde{k}_d\tilde{k}_f \geq 1$ and $\tilde{k}_f < 1$, decreasing χ_d results in agent 1 remaining at $\tilde{z}_1 = 1$ and agent 2 selecting a greater level of shading, which is good for the principal. For similar reasons as outlined above, when $\tilde{k}_d\tilde{k}_f < 1$, a decrease in χ_d may flip the $\tilde{k}_d\tilde{k}_f < 1$ inequality to $\tilde{k}_d\tilde{k}_f \geq 1$, resulting in agents changing from setting $\tilde{z}_1 = \tilde{z}_2 = 0$ to $\tilde{z}_1 = 1$ and $\tilde{z}_2 > 0$, which is also good for the principal.

In contrast to the results on χ_d, the closer agent 2's ideal point is to the set of actions that the principal wants the agent to conduct (smaller χ_f), the better the principal does. Taken together, the comparative statics on χ_d and χ_f suggest that, conditional on $\tilde{k}_d\tilde{k}_f \geq 1$, heterogeneous teams are most effective for the principal when the values of $|\chi_d|$ and $|\chi_f|$ are similar. Essentially, the principal does best when the agents' ideal points are symmetric around the principal's expected ideal point.

There is empirical evidence that insurgent leadership seeks agents with symmetric and offsetting preferences. In ISIS, insurgent leadership likely had little control over the preferences of foreign fighters, religious extremists who traveled great distances to fight to establish an Islamic caliphate. Given the extreme preferences of the foreign fighters, the theory predicts that ISIS's leadership should recruit extreme domestic agents whose preferences run counter to those of the extreme foreign agents. ISIS accomplished this by recruiting former members of the Arab Socialist Ba'ath Party, whose members historically possessed a strong political ideology rather than a religious identity (Fishman, 2016). ISIS is not the only group that started as a combination of individuals possessing distinct identities. The Red Commandos, a violent criminal organization that operates in Brazilian favelas, originated when leftist guerrillas joined forces with robbers and murders during their shared time occupying high-security prisons in the 1970s-1980s (Grillo, 2016, 29-43).

I include a further discussion of comparative statics in the Online Appendix.

4 How the Principal Behaves

Here I describe the principal's decision making. For this analysis, I assume that all agents want to be in the militant group. This assumption means that the worst equilibrium outcome for the agents in the second stage is still better than their reservation utility. I will relax Assumption 2 in the "Other Extensions" section.

Assumption 2: $-\left[\chi_f - \chi_d + \gamma\chi_f\right]/(1-\delta) \geq R_a$.[19]

4.1 Techniques to Prevent Subversion

I consider four techniques that a principal can employ in an effort to prevent subversion by agents. The principal can let the team operate independently (Hands-Off), form a heterogeneous team of agents (Heterogeneous Team), offer the smallest amount needed to completely align incentives (Incentive Contracts), or make a heterogeneous team and offer an optimal incentive contract (Heterogeneous Team with Incentive Contracts).

Definition: In the **Hands-Off Technique**, the principal does not issue transfers ($G_1 = G_2 = 0$), monitor the agents' per-period actions ($m = 0$), or designate team composition ($o_P = u$).

Definition: In the **Heterogeneous Teams Technique**, the principal forms a heterogeneous team ($o_p = f$), but does not issue transfers ($G_1 = G_2 = 0$) or monitor the agents' per-period actions ($m = 0$).

Definition: In the **Incentive Contracts Technique**, the principal monitors the agents' per-period actions ($m = 1$) and offers each agent transfers of $(1-\gamma)(a_{i,t} - \chi_d)$, but does not designate team composition ($o_P = u$).

Definition: In the **Heterogeneous Teams with Incentive Contracts Technique**, the principal sets $m = 1$, $o_p = f$, and offers transfers $G_{1,t}(a_1) = \hat{g}_1^*\left(a_{1,t,} - \chi_d\right)$ and $G_{2,t}(a_2) = \hat{g}_2^*\left(\chi_f - a_{2,t,}\right)$ where $\hat{g}_1^*$ and $\hat{g}_2^*$ maximize the principal's expected utility.

Within each technique, I describe how the agents play the game, I calculate the utilities, and I analyze comparative statics. I then discuss the principal's optimal choice across techniques. To preface what is to come, Figure 2 illustrates how varying monitoring

[19] The left hand side is the expected value a foreign agent would receive from being on a team with a domestic agent where both agents match actions to their ideal points.

costs κ_m and organizational costs κ_o influence which technique the principal selects, as well as examples of organizations that use the techniques.

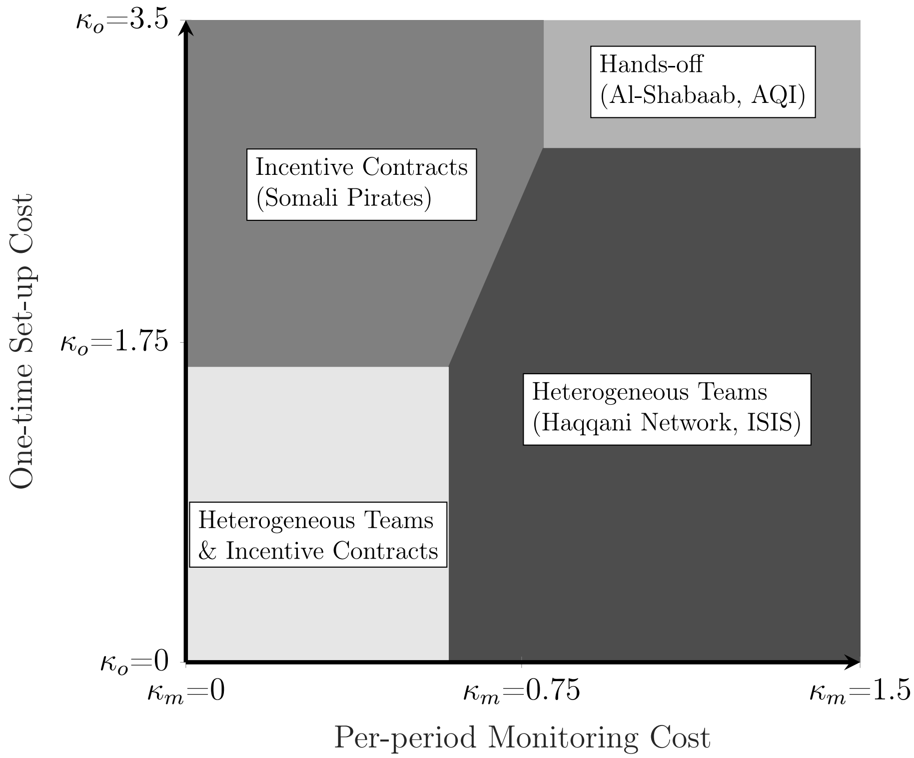

Figure 2: Optimal Team Management Techniques

Notes: Parameter values are $\gamma = 0.4$, $\delta = 0.8$, $\chi_d = -1$, and $\chi_f = 2.7$. For these parameters, optimal contracting with the Heterogeneous Teams with Incentive Contracts Technique was identified computationally using the Nealder-Mead simplex algorithm.

4.2 Heterogeneous Teams Technique

When the principal forms a heterogeneous team, agents will sometimes self-manage, which provides efficiency gains for the principal. How the agents behave was characterized in the definitions of $\tilde{z}_1$ and $\tilde{z}_2$, and comparative statics were explored in Observation 1. I summarize the actions and expected utilities in Proposition 1.

Proposition 1: When the principal employs the Heterogeneous Teams Technique:

- Agents set $a_{1,t} = (1 - \tilde{z}_1)\chi_d + \tilde{z}_1\omega_t$ and $a_{2,t} = \tilde{z}_2\omega_t + (1 - \tilde{z}_2)\chi_f$ for all $t \in \{1, 2, 3, ...\}$,
- $\mathbb{E}U_p = ((1 - \tilde{z}_1)\chi_d - (1 - \tilde{z}_2)\chi_f) / (1 - \delta) - \kappa_o$,
- $\mathbb{E}U_1 = (\tilde{z}_1\chi_d - ((1 - \tilde{z}_2)\chi_f - \chi_d) - \gamma(1 - \tilde{z}_1)(-\chi_d)) / (1 - \delta)$,
- $\mathbb{E}U_2 = (-\tilde{z}_2\chi_f - (\chi_f - (1 - \tilde{z}_1)\chi_d) - \gamma(1 - \tilde{z}_2)(\chi_f)) / (1 - \delta)$.

4.3 Hands-Off Technique

When the principal lets the team operate independently, agent 1 will form a homogeneous team, and the agents will match their actions to the state of the world and subvert.

Proposition 2: When the principal employs the Hands-Off Technique:

- Agents set $o_a = d$, $a_{1,t} = \chi_d$, and $a_{2,t} = \chi_d$ ($z_1 = 0$ and $z_2 = 0$) for all $t \in \{1, 2, 3, ...\}$,
- $\mathbb{E}U_p = 2\chi_d/(1 - \delta)$,
- $\mathbb{E}U_1 = \mathbb{E}U_2 = \gamma\chi_d/(1 - \delta)$.

Observation 2: When the principal employs the Hands-Off Technique, the principal's expected utility does not change with γ. The principal's expected utility is increasing in χ_d and is unchanging in χ_f.

The Hands-Off Technique results in some standard ally-principle type results. As long as $1 > \gamma$ (as assumed by Assumption 0), agents want to subvert. Thus, the closer χ_d is to the principal's expected most-preferred action ($\mathbb{E}(\omega_t) = 0$), the better the principal will do.

4.4 Incentive Contracts Technique

When the principal offers agents $G_{i,t} = (1 - \gamma)(a_{i,t} - \chi_d)$, agent 1 will partner with a domestic agent, and agents will not subvert.

Proposition 3: When the principal employs the Incentive Contracts Technique:

- Agents set $o_a = d$, $a_{1,t} = \omega_t$, and $a_{2,t} = \omega_t$ ($z_1 = 1$ and $z_2 = 1$) for all $t \in \{1, 2, 3, ...\}$,
- $\mathbb{E}U_p = \left(2\chi_d(1-\gamma) - \kappa_m\right)/(1-\delta)$,
- $\mathbb{E}U_1 = \mathbb{E}U_2 = (1+\gamma)\chi_d/(1-\delta)$.

Under the transfers defined above, two domestic type agents are indifferent over shading levels $z_i \in [0,1]$. Here agent 1 does best by selecting a domestic type partner, and the principal achieves the first-best contracting outcome where agents do not subvert.

Observation 3: When the principal employs the Incentive Contracting Technique, the principal's expected utility is strictly increasing in γ. The principal's expected utility is strictly increasing in χ_d and is unchanging in χ_f.

Like the Hands Off Technique, the Incentive Contracts Technique also results in ally principle results. As χ_d increases and γ increases, which means the agents' preferences are closer to those of the principal, it is less costly to buy good behavior through utility transfers.

4.5 Heterogeneous Teams with Incentive Contracts Technique

When the principal forms a heterogeneous team, oversees the agents' actions, and offers agents an optimal incentive contract, agents may shade their actions towards those that match the state of the world, which is what the principal wants. I limit my analysis to incentive contracts that adopt the common form of rewarding agents for deviating from their ideal points. Formally, the principal will select some optimal "transfer constants" $\hat{g}_1^*$ and $\hat{g}_2^*$ that define transfers $G_{1,t}(a_1) = \hat{g}_1^*\left(a_{1,t,} - \chi_d\right)$ and $G_{2,t}(a_2) = \hat{g}_2^*\left(\chi_f - a_{2,t,}\right)$.

I discuss the mathematical details of this technique in the Appendix. This includes Proposition 4 and Observation 4. The key results are that the principal's expected utility is weakly increasing in γ.

4.6 Management Costs and the Principal's Problem

Management costs influence the principal's technique selection. This model captures two types of management costs: κ_o denotes the one-time cost to oversee the organization of the team, and κ_m denotes per-period costs for monitoring the actions of the agents. The principal's choice of technique depends on the magnitude of these two costs. When the principal has low organizational costs κ_o and low per-period monitoring costs κ_m, the Heterogeneous Teams with Incentive Contracts technique generally is the most appealing to the principal.[20] If organizational costs increase but per-period monitoring costs do not increase, the principal may prefer Incentive Contracts. Alternatively, if per-period monitoring costs are high but organizational costs are low, then the principal may prefer Heterogeneous Teams. If all oversight costs are high, then the principal may prefer the Hands-Off Technique. This intuition is depicted in Figure 2.

In any setting, getting to know the agents' types, monitoring the agents' actions, and designing incentives to better motivate agents requires time and effort on the part of the principal. In an insurgency, this becomes an even greater challenge because any time the principal interacts with agents, the principal runs some risk of exposing the agents, being captured, or being killed. To offer more details, for insurgent groups, the costs κ_m of monitoring the actions of agents vary based on where the team of agents is operating. For example, for members of AQI operating throughout Iraq and members of the Haqqani Network operating in Afghanistan, insurgent leadership was under high pressure due to counterinsurgent targeting operations on the group's leadership; this would make monitoring agents in the conflict theater very risky (Schram, 2019). Alternatively, if the team of insurgents is operating within a zone of strong control (as discussed in Kalyvas (2006)) or in a "conventional" civil war (characterized by clear frontlines and an ability to deploy heavy weaponry (Kalyvas, 2005; Kalyvas and Balcells, 2010)), then the costs would be less. Values of κ_o would also rely on the factors above, but also on the degree to which the group's leadership enjoys a safe haven from which to operate. An insurgent group

[20] I say this is generally most appealing because, under some parameters, the principal can achieve full efficiency from the agents using the Heterogeneous Teams technique.

with a safe haven should have a smaller κ_o than an insurgent group without a safe haven, because the principal could safely organize the team in the safe haven and then send it into the conflict theater.[21]

There is evidence that the Haqqani Network (from the 1980s to the present) and ISIS (circa 2015) used a version of self-managing teams. Both groups benefited from the existence of a safe haven—the Haqqanis in Pakistan and ISIS in Syria—that reduced organizational oversight costs κ_o. And, both groups took pains to integrate foreign fighters into their operations. The leadership of the Haqqani Network created heterogenous teams of fighters and embedded small teams of foreign fighters alongside their regular forces (Brown and Rassler, 2013; Hamid and Farrall, 2015). Interviews in Weiss (2015) describe how ISIS initially created homogeneous, single-ethnicity battalions of foreign fighters, but found the battalions were more loyal to their leader than to the organization. In response, the organization called upon foreign, Syrian, and Iraqi fighters to integrate into battalions "through laying aside [their] prior identity for the muhajir and making him [the fighter] a resident in the Islamic State" (Al-Tamimi, 2015b,a). This is not to say that foreign and domestic foreign agents got along. Foreign and domestic fighters regularly clashed for ideological reasons (Brown and Rassler, 2013). However, both ISIS and the Haqqani Network were effective insurgencies despite introducing ideological extremists (Lilleby, 2013; Gates and Podder, 2015).

In contrast, AQI's leadership lost its safe haven in Iraq beginning in 2007, when US-led Coalition forces in Iraq captured or killed dozens of AQI's leaders (Schram, 2019). Following these operations, AQI embraced the Hands-Off Technique. Consistent with the model's equilibrium, domestic mid-level leadership began excluding foreign fighters from operations (CTC, 2007c). These homogenous teams then undertook a range of behaviors consistent with subversion (Fishman, 2009). For example, as AQI's leadership

[21] Admittedly, treating organizing as a one-off cost is likely an underestimation; a principal may initially form a heterogeneous team, but once the team is operational, some teammates may undermine the team structure. In practice, dictating organizational structure likely comes with more than a one-off cost, but would require less involvement than monitoring the day-to-day actions of the teams.

was reaching out to form an alliance with Ansar al-Sunnah, low-level members of AQI were killing members of Ansar al-Sunnah (CTC, 2007c,b). As a consequence of AQI members' violent and criminal behavior, local Sunni actors turned against the group and joined the "Awakening" movement instead, which played a major role in AQI's decline between 2007-2010 (Biddle *et al.*, 2012). However, by pursuing a Hands-Off strategy, enough of AQI's leadership was able to survive to reorganize as ISIS, which emerged as a serious insurgent threat in 2014 (Weiss, 2015, 114-130).

The contrast between the Haqqani Network, ISIS, and AQI is not only consistent with the comparative statics of the model, but it also illustrates how this model is a better fit of the case than alternate existing hypotheses on why polarized teams perform well. Other explanations for why polarized teams could outperform homogeneous teams include that different types of agents are willing to undertake tasks that their teammates are unwilling to undertake (Lazear, 1999), can offer new ideas on how to better conduct operations (Hong and Page, 2001), or can help motivate teammates and resolve some of the problems associated with free-riding (Baccara and Yariv, 2016); all of these are explanations that would suggest that agents (and not just the principal) benefit from being on polarized teams. However, across AQI and the Haqqani Network, foreign fighters and local fighters were at-odds and sought to exclude outgroup members when possible (or even killed outgroup members, in the case of al Shabaab), suggesting that polarized teams were undesirable for agents but desired by leadership (Fishman, 2009; Brown and Rassler, 2013; Hamid and Farrall, 2015).

To the best of my knowledge, there is no record of insurgent groups using Incentive Contracts to resolve subversion, with Bahney *et al.* (2013) investigating this question explicitly and finding that AQI did not use this technique. The avoidance of incentive contracts could be the consequence of the per-period monitoring that is required, as the model explores. There are other possibilities as well: implementing precise transfers to agents in organizations where grifting is common could be difficult (Shapiro and Siegel, 2012). While insurgencies do not seem to use incentive contracts, groups engaged in other illicit activities do. For example, groups that fund Somali pirates offer bonuses to agents who undertake risky tasks —like a $10,000 bonus for being the first person to board the ship that is being hijacked (Blanc, 2013).

It is not clear if any insurgent groups use the Heterogeneous Teams with Incentive Contracts Technique. It is possible to imagine a new or small militant group that is large enough to merit the principal-agents treatment while still being small enough that the principal can easily monitor agents' activities, provide flexible transfers, know the recruits' types, and change the organizational structure at will. In the licit sector, this is more common. For instance, new companies and sports teams often both dictate organizational structure as well as offer performance-based incentives.

5 Considering the "Perfectly Aligned" Agent

A natural intuition is that the principal could form a better self-managing team if one of the misaligned agents were replaced by a subordinate whose preferences were fully aligned with the preferences of the principal. This intuition does not always hold. Put another way, if the principal has a choice between an agent who values exactly what the principal values and an extremist foreign fighter to act as a teammate to a domestic fighter, in many cases, the principal can do strictly better selecting the foreign fighter. Why? Creating a team of a domestic agent and a perfectly aligned agent removes much of the useful strategic tension that exists between foreign and domestic teammates. Adding a perfectly aligned agent may be valuable, but its value is derived largely from ally-principle type results rather than from the strategic interactions between teammates.

I will consider a "perfectly aligned" agent, who has utility function

$$U_{pa} = \sum_{t=1}^{\infty} \delta^{t-1} \left(-(1+\gamma)|a_{pa,t} - \omega_t| - |a_{j,t} - \omega_t| \right),$$

and will select actions $a_{pa,t} = (1 - z_{pa})\omega_t + z_{pa}\chi_d$. When $z_{pa} = 0$, the perfectly aligned agent is selecting their most preferred action (which is also the principal's most preferred action), and when $z_{pa} > 0$, the perfectly aligned agent is selecting actions closer to their domestic partner's ideal point. In equilibrium, a perfectly aligned agent and a domestic agent will set $\check{z}_1$ and $\check{z}_{pa}$.

Definition: $\check{z}_1$ and $\check{z}_{pa}$ are defined as

- $\check{z}_1 = 0$ and $\check{z}_{pa} = 0$ if $\check{k}_d \check{k}_{pa} < 1$ and
- $\check{z}_1 = 1$ and $\check{z}_{pa} = \frac{1}{\check{k}_d}$ if $\check{k}_d \check{k}_{pa} \geq 1$,

where $\check{k}_d = \frac{-\delta\chi_d}{(1-\gamma)(1-\chi_d-\delta)}$ and $\check{k}_{pa} = \frac{-\chi_d\delta}{(1+\gamma)(1-\delta-\chi_d)}$.

Agent 1 is willing to shade up to $z_1 \leq min\left\{1, z_{pa}\check{k}_d\right\}$, and the perfectly aligned agent is willing to shade up to $z_{pa} \leq min\left\{1, z_1\check{k}_{pa}\right\}$. Because each agent's level of shading is an increasing function of their teammate's level of shading, when the perfectly aligned agent selects $z_{pa} > 0$, it can induce agent 1 to select an action that is closer to the principal's ideal point to an extent that may outweigh the disutility that the principal receives from $z_{pa} > 0$. Thus, selecting $z_{pa} > 0$ can follow from the maximization criterion on Assumption 1, and this occurs when $\frac{-\delta\chi_d}{(1-\gamma)(1-\chi_d-\delta)} > 1$ holds,[22] which always holds under the conditions in the second bullet point. I can then define equilibrium behavior and the principal's payoffs in Proposition 5.

Proposition 5: Assume the principal forms a heterogeneous team with one domestic and one perfectly aligned agent.

- Agents set $a_{1,t} = \check{z}_1\omega_t + (1 - \check{z}_1)\chi_d$ and $a_{pa,t} = (1 - \check{z}_{pa})\omega_t + \check{z}_{pa}\chi_d$ for all t,
- $\mathbb{E}U_p = \left((1 - \check{z}_1)\chi_d - \check{z}_{pa}(\omega_t - \chi_d)\right)/(1 - \delta) - \kappa_o$.

To compare the foreign-domestic team to the domestic-perfectly-aligned team (comparing

[22]This condition is derived in the Appendix and follows from taking first order conditions of the principal's utility function with respect to agent 1's level of shading.

Proposition 1 to Proposition 5), I must consider three distinct cases. First, when $\tilde{k}_f \geq 1$ (with $\tilde{k}_f$ defined preceding Proposition 1), then the domestic-foreign team is fully self-managing with both agents setting $a_{i,t} = \omega_t$. When this occurs, the foreign-domestic team always outperforms the domestic-perfectly-aligned team, which never sets $a_{1,t} = a_{2,t} = \omega_t$. Second, when $\tilde{k}_d\tilde{k}_f < 1$, then there is no productive shading occurring within the domestic-foreign team, meaning that replacing a foreign agent with a perfectly aligned agent will produce efficiency gains through ally-principle type results. Finally, when $\tilde{k}_d\tilde{k}_f \geq 1$ and $\tilde{k}_f < 1$, sometimes the domestic-perfectly-aligned team outperforms the domestic-foreign team, while at other times it does not.

6 Considering Averaging the Agents' Actions

The model above treated each agent's individual deviation from the state of the world as detrimental to the principal and the agent. An alternate setup benchmarks the average of the two agents' actions against the state of the world. The model above has the feature where if two agents subvert in different directions, two wrongs cannot make a right—in other words, agents selecting $a_{1,t} = -1$ and $a_{2,t} = 1$ is not the same as agents selecting $a_{1,t} = a_{2,t} = 0$ when $\omega_t = 0$. Because this feature may not hold in all cases, it is worthwhile considering if the logic of the shading equilibria above are robust to an alternate theoretical framework with averaged actions. To summarize what occurs in this setting, while moving from a homogeneous team to a diverse team produces a new, mechanical benefit for the principal, the principal stands to benefit further when agents strategically shade towards the state of the world.

In this extension (where I will not consider transfers), assume that each agent has utility function

$$U_i = \sum_{t=1}^{\infty} \delta^{t-1} \left(-|a_{i,t} - \chi_\tau| - |a_{j,t} - \chi_\tau| - \gamma | \frac{a_{i,t} + a_{j,t}}{2} - \omega_t | \right)$$

and the principal has utility function

$$U_p = \sum_{t=1}^{\infty} \delta^{t-1} \left(-|\frac{a_{i,t} + a_{j,t}}{2} - \omega_t| \right) - 1_{o \in \{d,f\}} \kappa_o.$$

For ease, I will assume $\chi_d = -1$ and $\chi_f = 1$, and realizations of ω_t are uniformly distributed on $[-1, 1]$. For agents to want to subvert when on a homogeneous team, I assume $\gamma < 2$.

I consider one possible shading equilibrium here. I define the following shading terms:

Definition: $\bar{z}_1$ and $\bar{z}_2$ are defined as

- $\bar{z}_1 = 0$ and $\bar{z}_2 = 0$ if $\bar{k}_d \bar{k}_f < 1$ and
- $\bar{z}_1 = 1$ and $\bar{z}_2 = 1$ if $\bar{k}_d \bar{k}_f \geq 1$,

where $\bar{k}_d = \bar{k}_f = \frac{\delta(2+\gamma)}{(4-3\delta)(2-\gamma)}$.

In equilibrium, for every period t where $\omega_t \leq 0$, agents select $a_{1,t} = -1$ and $a_{2,t} = 2\bar{z}_2\omega_t + 1$. For every period t where $\omega_t > 0$, agents select $a_{1,t} = -1 + 2\bar{z}_1\omega_t$ and $a_{2,t} = 1$. Consider when $\bar{k}_d \bar{k}_f \geq 1$, which implies that $\bar{z}_1 = \bar{z}_2 = 1$. When this holds, whenever the state of the world is further from agent i's ideal point relative to agent j's ideal point, agent j will match their action to their ideal point, and agent i will select the action such that the average of $a_{i,t}$ and $a_{j,t}$ matches the state of the world. When $\bar{k}_d \bar{k}_f < 1$, neither agent shades in, and their average action is always 0.

I can then define equilibrium behavior and the principal's payoffs in Proposition 6.

Proposition 6: Assume the principal forms a heterogeneous team with one domestic and one foreign agent.

- If $\omega_t \leq 0$, agents set $a_{1,t} = -1$ and $a_{2,t} = 2\bar{z}_2\omega_t + 1$. If $\omega_t > 0$, agents set $a_{1,t} = -1 + 2\bar{z}_1\omega_t$ and $a_{2,t} = 1$.

- $\mathbb{E}U_p = -\left(\frac{1-\bar{z}_1}{2}\right)/(1-\delta) - \kappa$.

Before discussing heterogeneous teams, under this new setup the principal's final utility from selecting a homogeneous team of domestic agents is $-\frac{1}{1-\delta}$. With this in mind, the principal's utility from forming a heterogeneous team ($-\left(\frac{1-\bar{z}_1}{2}\right)/(1-\delta) - \kappa_o$) illustrates the two-part value to forming diverse teams here. First, without any strategic behavior on the part of the agents (i.e. when $\bar{z}_1 = \bar{z}_2 = 0$), the principal achieves an efficiency gain by switching from a homogeneous team to a diverse team (with final utility $-\left(\frac{1}{2}\right)/(1-\delta) - \kappa_o$). Why? Because $\mathbb{E}(\omega_t) = 0$ and ω_t is uniformly distributed, a homogeneous team with average action -1 will mechanically be worse for a principal than a heterogeneous team with average action 0.[23] However, there is still room for improvement on the part of the principal; when the $\bar{k}_d$ term is large enough, agents are willing to undertake the shading procedure described above, which will result in the agent's average action always matching the state of the world, granting the principal the final utility of $-\kappa_o$.

7 Other Extensions

Overall, the model suggests that, conditional on Assumptions 1 and 2 holding, introducing more extreme agents into a team can lead to less subversion. I now discuss several variants to the model to better understand the results and assess their robustness. I consider two alternatives to Assumption 1: the case where agents maximize their joint utility and a more generalized form of "shading." I also consider what occurs when the agents' reservation utilities bind. I find that the results still hold, and in some cases are strengthened.

7.1 Agents Maximize Joint Utility

A natural concern with Assumption 1 is that it selects an equilibrium that is particularly good for the principal. As a simple alternative, I consider the case in which agents maximize their team's per-period expected utility. This is equivalent to assuming that agents could side-contract to one another and not be concerned with hold-up problems. To evaluate how

[23] A homogeneous team's average action will, in expectation, be a distance of 1 from the state variable. A heterogeneous team's average action will, in expectation, be a distance of $\frac{1}{2}$ from the state variable.

this new equilibrium selection rule changes outcomes, I compare the Incentive Contracts Technique with a team of domestic agents to the Heterogeneous Teams Technique.[24] Relative to the payoffs under Assumption 1, when agents maximize their per-period utility, Heterogeneous Teams induce a greater degree of self-management across a larger parameter space, while Incentive Contracts to domestic agents become more costly for the principal.

Proposition 7: Assume that agents maximize their joint per-period utility:

- Within the Incentive Contracts Technique, for agents $i \in \{1, 2\}$ and $j \in \{1, 2\}$ with $i \neq j$, the Principal transfers $G_{i,t} = (1-\gamma)(a_{i,t} - \chi_d) + a_{j,t} - \chi_d$. Agents set $a_{i,t} = \omega_t$. The principal receives expected payoff $EU_p = (2\chi_d(2-\gamma) - \kappa_m)/(1-\delta)$.
- Within the Heterogeneous Teams Technique, agents select $a_{i,t} = \omega_t$ and the principal receives expected payoff $EU_p = -\kappa_o$.

Recall that for a heterogeneous team under Assumption 1, for any shading to occur, the condition $\tilde{k}_d \tilde{k}_f \geq 1$ must hold (as described in Proposition 1). For a heterogeneous team under the assumptions here, agents will always match their actions to the principal's ideal point. In other words, here the principal no longer needs to worry that $\tilde{k}_d \tilde{k}_f \geq 1$, and so agents maximizing their joint utilities will always generate a weakly more favorable degree of shading for the principal. Proposition 7 shows that even if agents are disregarding what the principal wants, by maximizing their joint utility, they will, under a broader parameter set, do precisely what is best for the principal.

In contrast, changing from Assumption 1 to the assumption that agents maximize their joint utility makes the Incentive Contracts Technique worse for principal. Comparing Proposition 3 to Proposition 7, here the principal must pay each agent i an additional $(a_{j,t} - \chi_d)$ to get agents to match their actions to the state of the world. While a transfer of $(1-\gamma)(a_{i,t} - \chi_d)$ will make agent i indifferent over any action $a_{i,t} \in [\chi_d, \omega_t]$, agent i can still benefit when their teammate selects action χ_d (relative to action ω_t). Here the additional

[24] By limiting analysis to incentive contracts to domestic agents, the principal avoids the case where, after a set of incentive contracts designed for domestic agents is set, agent 1 selects a foreign type agent to obtain a greater team utility.

$(a_{j,t} - \chi_d)$ transfer is necessary to make the team of agents jointly indifferent over any action $a_{i,t} \in [\chi_d, \omega_t]$. Thus, overall, making this change to Assumption 1 makes Heterogeneous Teams weakly better and Incentive Contracts more costly.

7.2 Expanded Agents' Action Sets

In my definition of a shading equilibrium, I limited $z_i \in [0, 1]$. As a simple alternative, in the Appendix I consider the case in which $z_i \geq 0$, meaning agents can now select actions beyond the state of the world relative to their ideal point while still following the shading structure as before. I will refer to shading levels $z_i > 1$ as "over-shading" because, *ceteris paribus*, they describes the case where agent i shades beyond the level that the principal most prefers (which is $z_i = 1$). To summarize what occurs, allowing $z_i \geq 0$ opens a new degree of freedom in the maximization criterion within Assumption 2, which can lead to better outcomes for the principal. In some cases, agent 1 (the domestic agent) can over-shade to induce agent 2 (the foreign agent) to select a greater degree of shading than what is possible when $z_i \in [0, 1]$.

7.3 Increased Agents' Reservation Utility

In the Appendix, I analyze a model where Assumption 2 does not hold; instead, the principal must pay a flat rate to induce agents to remain in the terror group. To summarize what occurs when the agents' reservation utility binds, sometimes the principal must offer larger transfers to agents under the Heterogeneous Teams strategy relative to the Incentive Contracts strategy. However, because transfers in the Heterogeneous Teams strategy can be flat-rate transfers that are not conditioned on the agents' actions, the principal avoids the per-period κ_m payment, which can make Heterogeneous Teams less expensive than Incentive Contracts.

8 Conclusion

The model described in this paper presents a simple intuition for how diverse teams of agents can self-manage to reduce subversion. When allowed to choose their teammates, agents exhibit homophily and will at times subvert to the mutual benefit of their like-minded teammates. In contrast, when the principal requires that agents with different preferences work together, agents suffer when their different-type teammates subvert. Thus on a diverse team, agents may find a mutually beneficial compromise by not subverting, which is what the principal wants. While organizing a diverse team is not always possible—sometimes management cannot feasibly reach out to agents to ensure diverse teams form—forming diverse teams can be a low-cost way to mitigate agency problems. And, as the analysis above shows, this result is robust to a variety of assumptions and modeling technologies.

Perhaps the most surprising result to emerge from this model is that the principal will not always seek out agents with ideal points that are closest to the principal's own ideal point. As described in Observation 1 and explored in the "Perfectly Aligned Agent" example, the principal can achieve efficiency gains by utilizing fringe agents that offset the preferences of other agents within the organization. This result differs from the standard prediction in the ally-principle literature. What emerges is a more complex picture of what type of agent principals will prefer. In this decision, how agents interact within the team matters.

My analysis suggests several avenues for future work. One possibility is to analyze how the principal can allocate funds to specific agents to make their actions have more impact and what effect this has on the efficiency of self-managing teams. Another is to consider how the principal may experiment with team composition in settings where agents possess unknown utility functions.

References

Adams, R. B., Hermalin, B. E. and Weisbach, M. S. (2010) The role of boards of directors in corporate governance: A conceptual framework and survey, *Journal of economic literature*, **48**, 58–107.

Al-Tamimi, A. J. (2015a) The isis papers: a masterplan for consolidating power.

Al-Tamimi, A. J. (2015b) The islamic state 'masterplan' of administration- some analytical notes.

Baccara, M. and Bar-Isaac, H. (2008) How to organize crime, *The Review of Economic Studies*, **75**, 1039–1067.

Baccara, M. and Yariv, L. (2016) Choosing peers: Homophily and polarization in groups, *Journal of Economic Theory*, **165**, 152–178.

Bahney, B. W., Iyengar, R. K., Johnston, P. B., Jung, D. F., Shapiro, J. N. and Shatz, H. J. (2013) Insurgent compensation: Evidence from iraq, *American Economic Review*, **103**, 518–22.

Bathala, C. T. and Rao, R. P. (1995) The determinants of board composition: An agency theory perspective, *Managerial and decision economics*, **16**, 59–69.

Battaglini, M. (2002) Multiple referrals and multidimensional cheap talk, *Econometrica*, pp. 1379–1401.

Bendor, J., Glazer, A. and Hammond, T. (2001) Theories of delegation, *Annual review of political science*, **4**, 235–269.

Bendor, J. and Meirowitz, A. (2004) Spatial models of delegation, *American Political Science Review*, pp. 293–310.

Berman, E. (2009) *Religious, Radical, Violent*, The MIT Press.

Berman, E. and Laitin, D. D. (2008) Religion, Terrorism and Public Goods: Testing the club model, *Journal of Public Economics*, **92**, 1942–1967.

Bertelli, A. and Feldmann, S. E. (2007) Strategic appointments, *Journal of Public Administration Research and Theory*, **17**, 19–38.

Beyerlein, M. M. and Johnson, D. A. (1994) *Advances in interdisciplinary studies of work teams, Vol. 1.*, Elsevier Science/JAI Press.

Biddle, S., Friedman, J. A. and Shapiro, J. N. (2012) Testing the Surge: Why Did Violence Decline in Iraq in 2007?, *International Security*, **37**, 7–40.

Blanc, J.-B. (2013) *The pirates of Somalia: Ending the threat, rebuilding a nation*, International Bank for Reconstruction and Development, The World Bank.

Bonatti, A. and Rantakari, H. (2016) The politics of compromise, *American Economic Review*, **106**, 229–59.

Brown, V. and Rassler, D. (2013) *Fountainhead of Jihad*, Hurst & Company, London.

Bubb, R. and Warren, P. L. (2014) Optimal agency bias and regulatory review, *The Journal of Legal Studies*, **43**, 95–135.

Caparini, M. and Hwang, J. (2019) Police reform in northern ireland: Achievements and future challenges, *Stockholm International Peace Research Institute*.

Chai, S. (1993) An organizational economics theory of antigovernment violence, *Comparative Politics*, pp. 99–110.

Che, Y. and Yoo, S. (2001) Optimal incentives for teams, *American Economic Review*, **91**, 525–541.

CTC (2007a) Analysis of the State of ISI, Tech. rep., Combating Terrorism Center, West Point, NY.

CTC (2007b) Letter from Abu Hamza to Ansar al-Sunnah Highlighting Divisions s between Abu Hamza and a member of al-Ansar, Tech. rep., Combating Terrorism Center, West Point, NY.

CTC (2007c) Letters to the Leadership of ISI about Issues of Single-party Control, Tech. rep., Combating Terrorism Center, West Point, NY.

Dewatripont, M. and Tirole, J. (1999) Advocates, *Journal of political economy*, **107**, 1–39.

Dressler, J. A. (2010) *The Haqqani Network: From Pakistan to Afghanistan*, Institute for the Study of War.

Enders, W. and Jindapon, P. (2010) Network Externalities and the Structure of Terror Networks, *Journal of Conflict Resolution*, **54**, 262–280.

Felter, J. and Fishman, B. (2007) Al-qa'ida's foreign fighters in iraq: A first look at the sinjar records, Tech. rep., DTIC Document.

Fishman, B. (2009) Dysfunction and decline: Lessons learned from inside al-qa'ida in iraq, Tech. rep., DTIC Document.

Fishman, B. (2016) *The Master Plan: ISIS, al Qaeda, and the Jihadi Strategy for Final Victory*, Yale University Press.

Gailmard, S. and Hammond, T. (2011) Intercameral bargaining and intracameral organization in legislatures, *The Journal of Politics*, **73**, 535–546.

Gates, S. (2002) Recruitment and allegiance the microfoundations of rebellion, *Journal of Conflict Resolution*, **46**, 111–130.

Gates, S. and Podder, S. (2015) Social media, recruitment, allegiance and the islamic state, *Perspectives on Terrorism*, **9**, 107–116.

Gibbons, R. and Roberts, J. (2012) *The handbook of organizational economics*, Princeton University Press.

Gilligan, T. W. and Krehbiel, K. (1989) Asymmetric information and legislative rules with a heterogeneous committee, *American Journal of Political Science*, pp. 459–490.

Grillo, I. (2016) *Gangster warlords: Drug dollars, killing fields, and the new politics of Latin America*, Bloomsbury Publishing.

Hafez, M. M. (2010) Tactics, takfir and anti-muslim violence, *Self-inflicted Wounds: Debates and Division in Al Qaeda and its Periphery*, pp. 19–44.

Hamid, M. and Farrall, L. (2015) *The Arabs at War in Afghanistan*, Hurst.

Hegghammer, T. (2010) The Rise of Muslim Foreign Fighters: Islam and the Globalization of Jihad, *International Security*, **35**, 53–94.

Hegghammer, T. (2013) Should I Stay or Should I Go? Explaining Variation in Western Jihadists' Choice between Domestic and Foreign Fighting, *American Political Science Review*, **107**, 1–15.

Hirsch, A. V. and Shotts, K. W. (2015) Competitive policy development, *The American Economic Review*, **105**, 1646–1664.

Holmstrom, B. (1982) Moral hazard in teams, *The Bell Journal of Economics*, pp. 324–340.

Holmstrom, B. and Milgrom, P. (1990) Regulating trade among agents, *Journal of Institutional and Theoretical Economics . . .* , **146**, 85–105.

Holmstrom, B. and Milgrom, P. (1991) Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design, *JL Econ. & Org.*, **7**, 24–52.

Hong, L. and Page, S. E. (2001) Problem solving by heterogeneous agents, *Journal of Economic Theory*, **97**, 123–163.

Jackson, M. O. and Wilkie, S. (2005) Endogenous games and mechanisms: Side payments among players, *The Review of Economic Studies*, **72**, 543–566.

Kalyvas, S. (2006) The logic of violence in civil war, *Cambridge Studies in Comparative Politics, Cambridge University Press,*.

Kalyvas, S. N. (2005) Warfare in civil wars: Stathis n. kalyvas, in *Rethinking the nature of war*, Routledge, pp. 99–119.

Kalyvas, S. N. and Balcells, L. (2010) International system and technologies of rebellion: How the end of the cold war shaped internal conflict, *American Political Science Review*, pp. 415–429.

Kirdar, M. (2011) *Al Qaeda in Iraq*, Center for Strategic and International Studies.

Lazear, E. P. (1999) Globalisation and the market for team-mates, *Economic Journal*, pp. C15–C40.

Lilleby, L. W. (2013) The haqqani network: Pursuing feuds under the guise of jihad?, *CTX Global Ecco*, **3**.

Maciag, M. (2015) Where police don't mirror communities and why it matters, *Governing.*

McChrystal, G. S. (2013) *My share of the task: A memoir*, Penguin.

Moghadam, A. and Fishman, B. (2010) Self-Inflicted Wounds: Debates and Divisions Within al-Qaeda and its Periphery, Tech. rep., Combating Terrorism Center, West Point, NY.

Pettersson, T. and Öberg, M. (2020) Organized violence, 1989–2019, *Journal of peace research*, **57**, 597–613.

Ramsey, C. H. and Robinson, L. O. (2015) President's task force on 21st century policing.

Scahill, J. (2015) The purge: How somalia's al shabaab turned against its own foreign fighters, *The Intercept*.

Schram, P. (2019) Managing insurgency, *Journal of Conflict Resolution*, p. 0022002719832963.

Segal, I. (1999) Contracting with externalities, *The Quarterly Journal of Economics*, **114**, 337–388.

Shapiro, J. (2019) Formal approaches to the study of terrorism, in *The Oxford Handbook of Terrorism* (Eds.) E. Chenoweth, R. English, A. Gofas and S. N. Kalyvas, Oxford University Press, Oxford, chap. 10, pp. 148–162.

Shapiro, J. N. (2013) *The Terrorists Dilemma: Managing Violent Covert Organizations*, Princeton University Press.

Shapiro, J. N. and Siegel, D. (2007) Underfunding in Terrorist Organizations, *International Studies Quarterly*, **51**, 405–429.

Shapiro, J. N. and Siegel, D. A. (2012) Moral Hazard, Discipline, and the Management of Terrorist Organizations, *World Politics*, **64**, 39–78.

Tirole, J. (1986) Hierarchies and bureaucracies: On the role of collusion in organizations, *JL Econ. & Org.*, **2**, 181–214.

Trotsky, L. (1971) 1905.

Walter, B. F. (2017) The new new civil wars, *Annual Review of Political Science*, **20**, 469–486.

Weinstein, J. M. (2006) *Inside rebellion: The politics of insurgent violence*, Cambridge University Press.

Weiss, M. (2015) How isis picks its suicide bombers, *The Daily Beast website*, **16**.

Whiteside, C. (2016) New masters of revolutionary warfare: The islamic state movement (2002-2016), *Perspectives on Terrorism*, **10**, 4–18.

Wilson, J. Q. (1989) *Bureaucracy: What government agencies do and why they do it*, Basic Books.

Yeatts, D. E. and Hyten, C. (1998) *High-performing self-managed work teams: A comparison of theory to practice*, Sage.

Zedong, M. (1938) Problems of war and strategy, *Selected Works*, **2**, 224.