From Social Alignment to Social Control: Reporting the Taliban in Afghanistan

Patrick Bergemann,[a] Austin L. Wright[b]

a) University of California, Irvine; b) University of Chicago

Abstract: In many settings, witnesses can report wrongdoing to internal authorities such as officials within an organization or to external authorities such as the police. We theorize this decision of where to report as rooted in the policing of group boundaries, as the use of different reporting channels symbolically affirms or disaffirms affiliation with different social categories. As such, both witnesses and other social actors have an interest in where witnesses report. We evaluate this theory using villagers' reporting of illegal Taliban activity in Afghanistan in 2017 and 2018, where witnesses could report externally (e.g., to the national police) or internally (e.g., to village elders). We show how responses to wrongdoing arose from the interaction between self and others' attitudes toward the Taliban, and we reveal how reporting can be simultaneously punitive for the wrongdoer and affiliative for the category to which the wrongdoer belongs.

Keywords: crime reporting; social control; crime; social alignment; boundary maintenance; Afghanistan

W HEN individuals observe wrongdoing, they often face two choices: whether or not to report the matter, and whom to report the matter to. In many settings, witnesses can report to internal authorities such as social control agents within an organization or to external authorities such as the police or other enforcement agencies. For example, observers of many types of workplace wrongdoing have the option of reporting internally to a manager or human resources officer or to external legal authorities such as the U.S. Federal Bureau of Investigation or Securities and Exchange Commission (Dworkin and Baucus 1998; Jeon 2017). Similarly, victims and observers of sexual assault on college campuses can report to either campus authorities or to police (Fisher et al. 2003; Orchowski, Meyer, and Gidycz 2009). In these and in other cases, individuals must decide which type of sanctioning authority to activate, if any at all.

Importantly, this decision of to whom to report has consequences: different authorities have different means of responding to and punishing the behavior, and different authorities may be more or less sympathetic to a complaint and more or less motivated to act on it (Miceli, Near, and Schwenk 1991). One of the reasons child sexual abuse has persisted within the Catholic church for decades is not that no one reported priests' criminal behavior but that behavior was mostly reported to authorities within the church hierarchy who refused to address the problem rather than to external authorities like the police (Boston Globe 2003; Podles 2008).

Although some scholars have addressed the question of what motivates individuals to report to different authorities—particularly in the domain of whistleblowing—this topic remains undertheorized (Dworkin and Baucus 1998; Mesmer-Magnus and Viswesvaran 2005; Miceli, Near, and Dworkin 2008). Research tends to focus

on individual characteristics of witnesses, such as job tenure (Dworkin and Baucus 1998), trust in management (Brown 2008), and how individualistic the person is (Nayir and Herzig 2012), or on organizational characteristics, such as formal (Barnett, Cochran, and Taylor 1993) and informal policies (Sims and Keenan 1998) and ethical culture (Kaptein 2011).

A focus on individual characteristics of the witness and organizational factors, however, misses the fact that reporting others is also a social decision, influenced by others in the environment (Bergemann and Aven 2022; Paul et al. 2014; Ruback, Greenberg, and Westcott 1984), and reporters can suffer retaliation from peers if they do not approve of the report being made (Cortina and Magley 2003; Miceli et al. 2008). Peers may prefer certain perpetrators to be reported internally, as penalties may be more lenient, while preferring others to be reported externally, where punishments may be more punitive.

In this article we take a broader approach than that taken in whistleblowing research, first defining our core question as a problem relevant to all groups that have formal authorities that make and enforce rules (Ellickson 1987)—including organizations, families, and certain communities and social associations—and that are nested within a larger state or political unit. It is within these environments broadly defined that the decision of whether to report internally or externally becomes salient. Second, we theorize this decision as a means of policing or maintaining group boundaries. Reporting someone internally serves as an act of affiliation between that person's salient category membership and the group, whereas reporting someone externally represents the opposite. Thus, the decision of where to report takes on a social meaning above and beyond the punitive consequences for any individual perpetrator.

We test this theory in a setting where individuals made consequential decisions about whether to report wrongdoing to internal or external authorities. Specifically, we use data from thousands of villages in Afghanistan regarding villagers' reporting of the Taliban in the period spanning 2017 to 2018. When individuals observed illegal Taliban activity, they faced a choice: report the behavior to external authorities such as the Afghan National Army or National Police, report to internal authorities such as local village elders, or do not report at all. In particular, we exploit variation in whether villagers viewed the Taliban positively or negatively to see how this affected the reporting channel. By situating individuals within their social context, we show how responses to wrongdoing arise from the interaction between self and others and how reporting can both be punitive for the offender and affiliative for the category to which the offender belongs.

This work provides theoretical and empirical insight into when different reporting channels are activated by observers of wrongdoing, revealing that the decision is grounded in fundamental questions of group boundaries. Whereas concerns about group boundaries have long been known to affect behavior (Erikson 1966; Langley et al. 2019), the way they shape reporting has gone unrecognized. Our results also highlight how attitudes toward different social categories—based on race, gender, or other characteristics—may lead to systematic inequalities in outcomes for perpetrators by routing them to different reporting channels. Even a relatively fair justice system may produce biased outcomes so long as different categories

of people are reported to different authorities. Finally, this work sheds light on reporting patterns in conflict zones (Shaver and Shapiro 2021), showing how social control is not just a function of power or intimidation (Kalyvas 2006) but is a social process influenced by the attitudes of individuals living in contested areas.

Formal Social Control and the Decision to Report

Social control represents "the organized ways in which society responds to behavior and people it regards as deviant, problematic, worrying, threatening, troublesome or undesirable in some way or another" (Cohen 1985:1). Formal social control involves hierarchies that make and enforce rules (Ellickson 1987), often through legitimate, centralized authorities empowered with adjudication and enforcement (Baldassarri and Grossman 2011; Palmer and Feldman 2018), and the sanctions handed down through formal social control are viewed as representing official group expressions (Meier 2018). Organizations such as schools, churches, and workplaces; communities with local means of dispute resolution; and families all have relatively formalized means of making and enforcing decisions in this way. In some cases, anyone higher in the hierarchy has the authority to investigate and adjudicate wrongdoing, such as a manager or teacher; in others, specific individuals are designated to deal with the wrongdoing, such as a human resources officer or a university's Title IX office. The broadest, most general form of formal social control—governmental social control—is mandated and implemented by the state and takes the form of laws, courts, and the police (Black 1976; Cohen 1985).

Sources of formal social control are often nested within one another, defined by the extent of their jurisdiction and the punishments available to them. Although all people are subject to the control of the state, many people are also subject to more local authorities (Ellickson 1987), and these different sources of social control constitute distinct processes that function independently of one another. Whereas reporting a crime to either a local police officer or a Federal Bureau of Investigation agent activates the social control apparatus of the state, reporting a crime to an organizational authority activates a separate system. Furthermore, whereas the state has the capacity to investigate and punish without constraint, local sanctioning power is often more limited: punishments typically involve the loss of organizational privileges or expulsion from the group (Hollinger and Clark 1982).[1]

Because local sources of authority are nested within the overarching social control of the state, witnesses of wrongdoing must often decide which sanctioning body to activate, if any at all. Being subject to different social control agents, however, is only relevant in certain situations. If a particular behavior is only considered a violation by one authority but not another, then there is no choice to make regarding to whom to report. For example, being late to work may violate company policy, but it is not a crime according to the laws of the United States. Thus, reporting such behavior to the police is not a valid option. Similarly, a participant in a criminal enterprise would not report the organization's illicit activities to someone within that enterprise, as those activities would not be considered wrongdoing according to the internal norms of the group.

However, there is a large set of behaviors that violate both rules of formal groups and legal codes, and witnesses have a choice of where to report: internally, to authorities within local groups or organizations; or externally, to representatives of the state. Victims of sexual assault on college campuses and parents of children abused by priests can decide whether to report to their schools and churches, respectively, or instead to the police. Various forms of white-collar crime, such as embezzlement or fraud, can be reported internally within a workplace or externally to government representatives.[2]

The choice of whether to report internally or externally is consequential, as different authorities may have differing willingness to act in response to reports of wrongdoing (Miceli et al. 1991). For example, organizations that feel pressure to maintain their status or reputation may suppress reports of wrongdoing within their boundaries in order to avoid group-level repercussions and stigma by association (Piazza and Jourdan 2018). This may lead to more lenient treatment of serious perpetrators, which can result in environments where criminal behavior is likely to persist (Zipay et al. 2021), or may result in the relocation of problematic individuals, which may simply shift the illegal behavior to a different context (Keenan 2013). Furthermore, due to the fact that internal authorities are more likely to have personal relationships with perpetrators, they may be more willing to take into account local contingencies or extenuating circumstances, whereas external authorities may be more impersonal and punitive in their approach. In either case, reporting to one authority versus another has implications for the deterrence and punishment of wrongdoing both locally and within the broader society.

Despite the decision of which channel to activate being a common dilemma, however, the majority of research on reporting decisions either focuses on reporting to one particular social control agent (such as the police) or does not distinguish between the different types available (e.g., Cheng, Bai, and Yang 2019; Dungan, Young, and Waytz 2019). Part of the lack of attention is likely due to the fact that several whistleblowing studies have found that a majority of individuals report internally and that those who report externally often do so only after failing to have the matter addressed internally (Jeon 2017; Miceli and Near 2002; Rehg et al. 2004). Thus, external reporting is viewed as a rare contingency, which only becomes salient when internal options have been exhausted. However, this view has been challenged by a number of studies that suggest that the prevalence of internal reporting is context specific (Park et al. 2008). For example, a survey of employees of government agencies found that 36 percent of respondents preferred to report externally, with the number increasing to 47 percent for local councils (Zipparo 1999). Another study found that more than half of surveyed teachers preferred to go to an external authority (Gökçe 2013), and about a third of college women are more likely to report sexual victimization to the police rather than to campus authorities (Fisher et al. 2003). The likelihood of activating external authorities vis-á-vis internal authorities has been found to vary by industry (Rothschild and Miethe 1999), employee level (Gao, Greenberg, and Wong-On-Wing 2015), and whether an organization is public or private (Nayir, Rehg, and Asa 2018). Put together, the dilemma of whether to report to an internal or external authority is not an uncommon one, and reporting externally is not simply a residual category.

Despite its perceived rarity, the question of to whom individuals report has been addressed to a limited extent in whistleblowing research, where the focus has been on individual and organizational factors (Dworkin and Baucus 1998; Jeon 2017; Mesmer-Magnus and Viswesvaran 2005; Miceli et al. 2008). Although some studies find no individual differences between internal and external reporters (Culiberg and Mihelič 2017; Jubb 1999), others have identified shorter tenure in the organization (Dworkin and Baucus 1998), less trust in management (Brown 2008), and how individualistic a person is (Nayir and Herzig 2012) as predictors of external reporting. At the organizational level, formal (Barnett et al. 1993) and informal policies (Sims and Keenan 1998) affect which reporting channel is used, as does ethical culture (Kaptein 2011).

Together, these investigations largely align with a prosocial view of whistle-blowing (Dozier and Miceli 1985), where reporting internally is seen as a means of showing commitment and loyalty toward the organization (Chen and Lai 2014; Donkin, Smith, and Brown 2008; Jeon 2017), as external reporting is perceived as endangering the organization due to the possibility of public embarrassment or loss of legitimacy (Lee and Fargher 2014; Vadera, Aguilera, and Caza 2009). Despite its contributions, however, this perspective—with its focus on individual and orga-nizational factors—has led researchers to neglect the role of social influence in the decision of where to report (Bergemann and Aven 2022). Reporting is known to be affected by others in the environment (Bergemann 2019; Paul et al. 2014; Ruback et al. 1984), and reporters can suffer retaliation if their peers do not approve of the report being made (Cortina and Magley 2003; Miceli et al. 2008). Presumably peers also respond to where wrongdoing is reported, that is, whether a report is made internally or externally. Furthermore, previous research has neglected how the identity of the perpetrator—particularly the perpetrator's social identity—may affect the authority to which wrongdoing is reported. Suggestively, Fisher et al. (2003) found that sexually victimized college students were more likely to go to the police when the perpetrator was a different race or ethnicity or was a stranger, whereas victims were more likely to go to campus authorities when they knew the perpetrator.[3] This suggests that features of the perpetrator may also be critical in determining to whom a witness or victim decides to report.

Social Categorization and the Activation of Different Authorities

Internal and external reporting do not simply represent different means of social control but also have relevant social implications. Because internal social control is practiced within a group, it generally applies to individuals who are members of or are affiliated with that group. Internal authorities must have some jurisdictional power in order to mandate and enforce judgments. Thus, reporting someone inter-nally can be said to have a symbolic meaning beyond its practical implications; it symbolically connects or affiliates the person reported with the group. Reporting internally tacitly acknowledges the internal authority's right to regulate the perpe-trator's behavior, whereas reporting externally disaffirms the perpetrator's right to

be treated as a member of the group. In the latter case, the perpetrator is reported to authorities who are explicitly outside the group, distancing the group from both the perpetrator and the investigative process.

Importantly, this symbolic affirmation or disaffirmation does not occur solely at the individual level, that is, only in regard to the individual perpetrator. Such acts also have categorical significance. Social categorization is a universal mechanism for making sense of the social world, whereby individuals are classified in culturally meaningful ways, such as by race, gender, nationality, and sexual orientation (Hogg 2004; Rhodes and Baron 2019). Of course, the same people can be categorized in multiple ways, as categorizations are diverse and overlapping (Crisp and Hewstone 2007), but certain categories tend to be more or less salient (Hogg and Turner 1987). For example, gender and race are salient in many situations, whereas other categorizations may be more context specific (Van Knippenberg and Dijksterhuis 2000).

Reporting internally implicitly affirms a perpetrator's salient category as being affiliated with the group, whereas reporting externally implicitly distances a perpetrator's salient category from the group.[4] To the extent that certain categories of perpetrators are systematically reported to internal or external authorities, the entire category becomes increasingly affiliated with or distanced from the group. For example, sexual victimization among college women is more likely to be reported externally to the police (as opposed to internally to campus authorities) when the offender is African American or Hispanic (Fisher et al. 2003). In the aggregate, this leads these minorities to be practically excluded from internal social control and symbolically disaffiliated from the college community.

These acts of affiliation and disaffiliation represent an unrecognized mechanism by which social boundaries are negotiated and can serve as sites of contestation and political struggle (Keblusek, Giles, and Maass 2017; Light and Iceland 2016). Individuals construct or reconstruct boundaries to either integrate or differentiate between actors, which establishes values about what is important to the group (Drori, Wrzesniewski, and Ellis 2013) and is a central mechanism in collective identity construction (Hernes 2004; Hunt and Benford 2004). As summarized by Wimmer (2007), "individuals and groups struggle over who should be allowed to categorize, which categories are to be used, which meanings they should imply and what consequences they should entail" (P. 11).

Thus, the decision of where to report has implications for which categories of people are perceived as more or less affiliated with the group, or with whom group members desire to be more or less affiliated. As such, the decision of where to report is of interest, not just to individual witnesses, but to others in the social environment. As the reporting channel has implications for social boundaries, we therefore expect it to be influenced—either implicitly or explicitly—by the views of others. Peer retaliation, which can occur when the group disapproves of a reporting decision (Rehg 1998), is likely to take place if an individual reports to a particular authority in contravention of the group's wishes.

We expect these dynamics around social boundaries to manifest themselves in the attitudes that witnesses and others in the social environment hold toward the salient category or categories to which a perpetrator belongs. Such attitudes

represent positive or negative views of certain social categories, and we expect the interaction between individual and social attitudes to influence whether wrongdoing is reported to internal authorities, to external authorities, or not at all.

We label the relative agreement or disagreement of attitudes toward social categories between witness and other social actors as *alignment*. It is the interaction of these attitudes that we expect to affect the decision of where to report a member of that category. *External alignment* occurs when witnesses and other group members have unfavorable views of a perpetrator's salient category, whereas *internal alignment* occurs when witnesses and other group members have favorable views of a perpetrator's salient category. Misalignment occurs when individual attitudes and social attitudes differ. Alignment indicates either a consistent or conflicted viewpoint of the wrongdoer's salient category, which should affect both whether the wrongdoing is reported and what social control agents are activated. The following describes each of the four quadrants that results from the interaction between these individual- and group-level attitudes.

Consider first the case of external alignment, where witnesses and other group members hold negative views of a perpetrator's salient category. Because there is consensus that this category should be disassociated from the group, the perpetrator's crimes would be appropriately investigated and punished by external authorities. In this way, the perpetrator—and indirectly the category—are excluded from group institutions and symbolically shunned from the group.

In the case of internal alignment, positive views toward the perpetrator's salient category lead witnesses and other group members to view the transgression as an internal matter. Witnesses will still want to report because they have an interest in preventing wrongdoing within the group and regulating the behavior of others, but they will primarily go to internal authorities. This allows them to symbolically acknowledge affiliation with the perpetrator's salient category while still attempting to control the transgresser's behavior. The group will similarly respond approvingly and will be less likely to retaliate; the witness is appropriately maintaining group boundaries.

This leaves cases when individual and group attitudes are misaligned. There are two types of misalignment, both of which likely result in a lower likelihood of reporting either internally or externally. The first is when a witness holds negative attitudes toward a wrongdoer's salient category while other group members hold positive attitudes. There are a couple of ways in which this might occur. The first is due to personal experience, where past negative interactions with members of that category may lead an individual to adopt negative attitudes, in contrast to other members of the group. Exposure to even a single negative incident can lead to stereotyping and prejudice against an entire category of people (Henderson-King and Nisbett 1996). Being the victim of wrongdoing may particularly spur this process, due to the personal nature of the violation. Alternatively, a victim may find a crime so egregious that they desire the individual perpetrator to face the harshest penalty possible, regardless of the perpetrator's category membership. In either case, reporting such a perpetrator internally might privilege the category in a way the witness finds undesirable, or be overly lenient for the perpetrator. And reporting externally may lead to retaliation or ostracism from other group members,

	Do not report	Report (External authorities)
Negative		
Positive	Report (Internal authorities)	Do not report

Individual Attitude toward Perpetrator Category

Positive   Negative

Group Attitudes toward Perpetrator Category

Figure 1: Model of reporting behavior based on individual and group attitudes toward a perpetrator's salient category.

given the symbolic implications of the act. Given these circumstances, a witness may be less likely to report at all.

The fourth and final case occurs when witnesses hold positive attitudes about a perpetrator's salient category while others hold negative attitudes. Again, misalignment indicates that witnesses are less likely to report. Reporting directly to external authorities would be disaffiliating from a category that the witness feels should be treated as a part of the group, escalating an issue that the witness likely believes should be handled internally. Yet local authorities are likely to treat perpetrators as if they were outsiders, perhaps by sharing the report with external authorities such as the police or punishing transgressors more harshly than is typical. This type of misalignment may be most likely to occur when members of stigmatized minorities witness wrongdoing by other minority members.

The theory is depicted in Figure 1, which shows the expectations of the model. Put more formally into hypotheses, we expect to observe the following:

Hypothesis 1 (H1): When individuals and other group members are aligned in attitudes toward a salient category, witnesses of wrongdoing committed by members of that category will be more likely to report.

Hypothesis 2 (H2): When individuals and other group members are aligned in negative attitudes toward a salient category, witnesses of wrongdoing committed by members of that category will tend to report to external authorities.

> **Hypothesis 3** (H3): When individuals and other group members are aligned in positive attitudes toward a salient category, witnesses of wrongdoing committed by members of that category will tend to report to internal authorities.

In order to test these hypotheses, we require data that not only have information about reporting behavior to internal and external authorities but also provide insight into how a perpetrator's salient category is viewed at both the individual and the group level. Furthermore, we need a large enough sample to allow us to assess the effects of these differences across groups. We use the results of a unique survey administered in 2017 and 2018 to individuals in thousands of villages across Afghanistan as to whether or not they observed and subsequently reported illegal Taliban activity. This setting has the benefit of involving a very salient category—the Taliban—and being a real-world context with real-world consequences (Fiorin 2021); vignette studies that ask about hypothetical intentions or lab experiments cannot capture the potential social, economic, and safety consequences that those deciding whether or not to report wrongdoing must confront (Cortina and Magley 2003; Miceli et al. 2008), which helps to explain why hypothetical intentions to report often have little relation to actual reporting behavior (Mesmer-Magnus and Viswesvaran 2005).

Social Control in Afghanistan

The recent history of Afghanistan is one of instability.[5] From various civil conflicts to the Soviet–Afghan War in the 1980s to the rise of the Taliban in 1996 and the invasion by the United States in 2001, the territory has rarely been able to rely on central authorities to maintain peace and order (Karell and Freedman 2019). Within this vacuum, villages have become supremely important as the center of social, economic, and administrative activity (Habib 2013). With more than 80 percent of the population residing in rural villages (Katzman 2011), the village "remains for most of its inhabitants the most significant institution in their lives and collective action at the village level will continue to have a primary role in ensuring the provision of public goods" (Pain 2016:17). Over time, local systems of internal social control—known as *jirgas* and *shuras*—have developed and maintain a prominent role in village life (Wardak and Braithwaite 2012).

Following traditional laws, jirgas represent dispute settlement for those of the Pashtun tribe, whereas shuras are commonly used for dispute settlement among other ethnic groups. Both operate similarly (Strand 2013) and generally involve an ad hoc group of respected individuals—primarily village elders and community leaders—who convene in order to adjudicate a dispute. Although these groups are ad hoc, they are not arbitrary; they represent tribal forms of dispute settlement whose members are relatively constant. As Alekseyeva et al. (2017:51) note, "Afghanistan's local adjudicatory systems share many of the 'formalities' of codified legal systems, and can rightly be viewed as institutions applying customary law." These institutions remain nested within the broader apparatus of the state, representing a local source of internal authority that differs from external authorities (e.g., police and military forces) (Coburn 2011; Murtazashvili 2016).

In order to activate these sources of non-state justice, villagers first approach elders to report grievances and seek mediation. These grievances encompass a wide range of behaviors, from minor problems such as defining property boundaries and minor bodily harm to major issues such as communal land disputes, murder, and crimes committed by the Taliban (Wardak, Saba, and Kazem 2007). Infractions committed by the Taliban include harassment, corruption, double taxation, and excessive use of force. Drawing on numerous interviews of village elders and Taliban commanders and fighters, Farrell and Giustozzi (2013) provide compelling accounts of the role elders play in mediating such grievances. Elders have several tools at their disposal when a grievance is brought to them, as decisions can incorporate a variety of social and economic sanctions. These include peacemaking and reconciliation, collective boycotts, compensation for victims, the marriage of a woman from an accused's family to a victim's close relative, and burning down the offender's house (Wardak et al. 2007). Elders can also use their social and political capital to channel concerns about Taliban activity directly to commanders in charge of offending rebel soldiers (Berman et al. 2011). Similar to other forms of internal social control, elders have limited and less dramatic tools at their disposal than government representatives.

Whereas village systems of social control maintain local jurisdiction, the national government in Afghanistan provides a separate, external source of social control. The system during our period of observation was put in place shortly after the American-led invasion of Afghanistan in 2001 and consisted of the police (local and national) and the national army. Although Afghan forces were weak and had outdated equipment at the time of the U.S.-led invasion, coalition forces provided massive amounts of training, equipment, and funds to develop the military and police into a more modernized force (Fetzer et al. 2021). The national army went from an "irregular militia" in 2001 to more than 160,000 soldiers in 2016 (Beal 2016). Although its overall effectiveness has rightly been called into question (Giustozzi 2015), the army and police maintained a presence throughout Afghanistan and were responsive to reports of illegal Taliban activity throughout our period of study in 2017 and 2018.[6] Their methods, as with governmental social control in general, were less constrained than those used by village elders and often involved the use of force and violence against insurgent forces.

The act of reporting to internal or external authorities therefore sought to accomplish the same objective: punishing and deterring wrongdoing within the village. Although neither channel was a guarantee of success, both activated social control agents who had the capacity to prevent and punish the illegal behavior. Reporting to elders represented a way of dealing with wrongdoing that kept knowledge of the incident local, whereas reporting to the police or the army informed and involved external forces. This distinction can be clearly seen in the story of an aggrieved woman who appeared at a wedding as recounted by Barfield (2013):

> An old woman burst into the men's celebration to demand that the notables there hear her complaint against a man in attendance who had borrowed 250 Afghanis—roughly equivalent to US$5—from her and refused to repay it [...] After quick whispered conference, the notables ordered the man to pay her on the spot [...] The notables then called

her back to warn her that she must consider the case resolved now. In
particular, they said she should not take her dispute to the government.
(P. 137)

Our focus is on understanding the conditions under which observers of wrong-
doing committed by members of the Taliban—a highly salient category—reported
internally or externally or did not report at all (Sonin and Wright 2022). In particular,
we seek to understand this behavior in terms of individual- and village-level dynam-
ics. To properly evaluate our hypotheses, however, we must hold constant aspects
of the broader political environment—particularly regarding political instability—
present in Afghanistan during our period of study when the Taliban had a presence
in 70 percent of the country (Sharifi and Adamou 2018). We therefore examine
variation *within* the 398 districts in Afghanistan, holding constant factors such as
state capacity and the extent of Taliban control, both of which are known to affect
reporting decisions (Kalyvas 2006). Given the prevailing conditions within a district,
we ask whether variation in individual and group attitudes toward the Taliban led
to the reporting pattern predicted in Figure 1.

Note that even with these controls, a relevant question is to what extent the
political context in Afghanistan nevertheless shaped behavior in ways that we
should not expect in other, non-combative settings where reports can be made
to internal and external authorities. As will be discussed in the results, very few
reports were made directly to the opposing forces of the conflict, that is, the Taliban,
the national army, or international forces. Instead, individuals overwhelmingly
reported to village elders or the national police, both of which served as civilian
sources of formal social control. Thus, we view this setting as comparable to other
situations where individuals decide between reporting to internal and external
authorities, albeit as an extreme case. However, extreme cases can facilitate theory
building by helping to make difficult-to-observe dynamics more visible (Eisenhardt
1989; Pettigrew 1995).

Data and Design

Data Collection

The data we use come from the U.S. Central Command, which contracted the
Afghan Center for Socio-economic and Opinion Research (ACSOR), an Afghan
subsidiary of the international firm D3, to design and field a recurring survey
(known as Project Foghorn) on the attitudes and experiences of individuals living
in Afghanistan. ACSOR hired and trained local enumerators in household and
respondent selection, how to correctly record answers to questions, culturally sen-
sitive interview methods, and secure storage of contact information (Child 2019).
ACSOR's use of local-to-area enumerators increases comfort with survey interviews
and decreases anxiety that external actors are monitoring and tracking respondents.
Respondents were not aware of the agency, organization, or government for which
the survey was commissioned, and ACSOR itself conducted surveys throughout the
study period for dozens of different agencies, nonprofit organizations, and govern-
ments (Child, Wright, and Xiao 2021; Condra and Wright 2019). We use waves 33,

34, and 35 of these quarterly surveys, which were collected in November/December 2017, February 2018, and May 2018. We study these waves because these are the only surveys that include precisely geo-referenced village locations.

The survey was stratified by province, covering all of the 34 first-level administrative units in Afghanistan. The administrative district was the primary sampling unit.[7] Sampled districts were selected via a probability proportional to size approach. After districts were selected for sampling, secondary sampling units composed of villages and settlements were randomly selected. After the sampling set had been identified and before fielding a survey wave, ACSOR engaged with local elders to secure permission for enumerators to enter sample villages. Once enumerators arrived at a village, a random walk method was used to identify target households. The random walk began from one of five randomly chosen locations (the center of the village or the northern, southern, eastern, or western border), and five households were surveyed within each village. Once a household was selected, a Kish grid was used to randomize the respondent within each selected household.

Survey diagnostics from other programs collected by ACSOR suggest that refusal, cooperation, and non-contact rates were low. We have the most systematic data on these diagnostics for the Afghanistan Nationwide Quarterly Assessment Research (ANQAR), which was collected by ACSOR in waves from a comparable time period (2012 to 2017). The survey sampling design and approach to question phrasing is shared between Foghorn and ANQAR, although the question topics vary between the two survey platforms. ACSOR's response rate exceeds 81 percent in all rounds. The high response rates suggest that respondents' decisions to engage with survey enumerators and discuss potential illegal activities conducted by the Taliban are unlikely to confound our study.

Although the high response rate of the survey is a positive for statistical inference, it raises questions about how the survey was conducted, especially if respondents somehow suspected that the United States was funding the survey. This may have led respondents to feel pressure to respond and may have led them to answer in a biased manner. To address this concern, we examined cooperation rates of another survey undertaken by ACSOR—the Survey of the Afghan People (SAP)—which is commissioned by the Asia Foundation and has no connection to global or local governments. The Asia Foundation is a nonprofit international development organization, and the SAP has been collected 14 times since 2004. The survey in 2018 had a similar response rate, suggesting that our surveys' connection to the U.S. government did not affect response rates and that these levels are typical of ACSOR surveys.

Instead, the high cooperation rates in both surveys are likely attributable to the fact that these surveys are conducted face-to-face by enumerators who are residents of the provinces in which they conduct interviews. Enumerators—who also do not know the original funding source—are trained to carefully manage access to villages and to be culturally sensitive at all times. Suggestive evidence for this comes from survey results for 2020, when COVID-19 forced ACSOR enumerators to switch to telephone-based survey methods. The survey data compiled during August and September of 2020 had a cooperation rate of only 44 percent. By comparison, there was only a two percent reduction in cooperation after the 2015

Hindu Kush earthquake—the most powerful seismic disaster in Afghan history—when interviews continued to be conducted face-to-face. We interpret this shift as an indication that the method of survey implementation (in-person versus over the phone) matters and that ACSOR's standard approach yields high levels of respondent engagement.

Furthermore, the survey responses themselves suggest that respondents did not feel pressure to respond in particular ways. For example, one question in the survey asks "Do you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable view of International Forces?" A majority (53 percent) of respondents indicated that they had a somewhat unfavorable (28 percent) or very unfavorable (24 percent) view of International Forces. If respondents felt coerced or were subject to social desirability bias, we would expect these numbers to be far lower. Furthermore, our survey and the 2018 SAP asked a similar question regarding attitudes toward the Taliban. In the SAP, 82.4 percent of respondents indicated "No sympathy at all" for the Taliban, which is similar to the 87.1 percent of respondents in our study who indicated somewhat unfavorable or very unfavorable views of the Taliban. Furthermore, the SAP found that five percent of respondents had "A lot of sympathy" for the Taliban, which is similar to the four percent of respondents we found who held very favorable views toward the Taliban. Together, these pieces of evidence make us confident in the general validity of the survey used for the analysis. To further validate data collection, ACSOR has set up a battery of audits conducted randomly to evaluate field operations and enumerator activity. Taken together, these diagnostic trends and measurement and data safeguards give us confidence that Foghorn represents a highly useful survey platform for research, and ACSOR data have previously been used for academic research (e.g., Berman et al. 2011; Condra et al. 2018).

Measuring Group Categorization and Other Variables

We use subject responses in Foghorn to identify individuals who had been exposed to illegal activities carried out by the Taliban and evaluate their subsequent responses to questions regarding whether and to whom they reported. Overall, 25 percent of the 42,223 survey respondents indicated that they observed illegal Taliban activity in their villages in the previous month, as they answered affirmatively to the question "In the last month, have you seen the AGE [anti-government elements] doing something illegal in your community, or not?" These individuals constitute the sample that we use for our analysis. Note that the question specifically asks about *illegal* activity. In other words, the question is not whether respondents observed the Taliban at all, but specifically whether they observed illegal activity such as violence, intimidation, or double taxation. Our dependent variable—whether a villager exposed to illegal behavior reports it—comes from the following survey item: "Did you report any of the incident/s, or not?"

To measure attitudes toward the Taliban, we make use of a particular survey question that asks, "Do you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable view of the following groups?" One of the groups rated was the Taliban. Crucially, we can identify the location of villagers and link

them to other members of their community. This allows us to aggregate the attitudes of community members by village (excluding the respondents themselves). These aggregated preferences similarly range from strong support to strong opposition but are continuous because individual preferences are averaged. We are particularly interested in whether and how individual attitudes align with community attitudes. Conceptually, it is easiest to conceive of our approach as a two-by-two table (see Figure 1). When an individual supports the Taliban and their village's aggregate attitudes are similar, a particular respondent is in the lower-left quadrant. If an individual's attitudes align with their village in opposition to the Taliban, that respondent is in the upper-right quadrant. Off-diagonal quadrants (the upper left and lower right) are each cases of misalignment but with varying individual and village preferences for the Taliban. To capture this joint variation in individual and village views, we use an interaction term with constitutive base terms.

Public support for the Taliban varied geographically and was highest where the Taliban was most active in the southern and southeastern provinces. If the Taliban was more likely to break the law in communities where they were more active, the relationship between attitudes and reporting may be confounded and biased. To account for this spatial correlation in reporting and support, we include fixed effects for the 398 administrative districts in our benchmark model specification. This allows us to hold constant local power balances and state capacity while examining our hypotheses.[8]

Analogously, Taliban activity between survey waves may explain variation in public support. For example, on January 20, 2018, between waves 33 and 34, the Taliban conducted an attack on the Inter-Continental Hotel in Kabul. These types of attacks and any other factors that cause "common shocks" across the country potentially influence civilian (and village) views of the Taliban as well as the willingness to report wrongdoing. To account for these types of factors, we also include survey wave fixed effects in our benchmark specification.

The samples in our survey are redrawn each quarter, meaning that the data are not longitudinal. Because our sample composition changes in each wave, we incorporate individual demographic characteristics of respondents to address any concerns about the correlation between ethnicity, religion, age, marital status, gender, and education level (some of which have been found to affect reporting in other contexts [Mesmer-Magnus and Viswesvaran 2005; Miceli et al. 2008]). One natural concern, for example, is that relatively more educated individuals may have different attitudes or may behave differently than less educated community members. Likewise, age and wealth may account for some of the variation in whether individuals' attitudes align with their village.

The unique richness of our survey data enables us to account for other potential sources of omitted variable bias. Working respondents might suffer relatively more from armed group victimization than unemployed participants, perhaps causing them to report relatively more frequently. Similarly, elders, individuals personally victimized by wrongdoing, participants exposed to different types of crimes, and those with differing views of governmental social control agents might respond differently when witnessing illegal behavior. Although these respondent characteristics are not the primary quantity of interest in our study, incorporating

them in our various model specifications increases confidence in the conditional independence assumption when interpreting our main effects.

Besides omitted variable bias, other potential biases are also important to address. This is particularly true given the sensitive nature of some of the questions asked in this survey, which are used to construct our key variables. To address concerns about additional biases—for example, regarding exposure to wrongdoing, interpretation of behavior as wrongdoing, and social desirability bias—we present a variety of supplemental tests subsequent to the main analysis.

Benchmark Empirical Specification

Our empirical design leverages the structure of our data: we observe individual-level responses and can identify the villages within which each respondent is embedded. This allows us to measure each individual's support as well as to aggregate the preferences of their village at the same time. Our approach to the quantity of interest—social alignment—is to use an interaction term that captures the marginal effects of simultaneous variation in support for the Taliban along both individual and village dimensions. Because our outcome variable is binary, we introduce results from a linear probability model. However, note that all results have been replicated using logistic regressions and are substantively the same. Our regression-based estimates enable us to assess the statistical significance of each parameter, although interpreting the marginal effects of the ordinal measures is less straightforward. If our results are consistent with our theoretical argument, we would expect that the probability of reporting should be decreasing in each of the base terms (negative coefficients) and increasing in the interaction (positive coefficient). To ease the substantive interpretation of the magnitudes of these effects, we construct contour plots for each of our main model specifications. Our theory suggests that we should expect reporting to be highest when individual preferences converge with the community at either jointly high or jointly low levels of support for the Taliban.

We begin by studying Equation (1):

$$
\begin{aligned}
y_i = \alpha &+ \beta_1 Individual_i + \beta_2 Village_i \\
&+ \beta_3 Individual_i \times Village_i \\
&+ \lambda D_i + \zeta W_t + \gamma X_i + \epsilon,
\end{aligned}
\tag{1}
$$

where y_i indicates whether the respondent reported the illegal wrongdoing they witnessed the Taliban commit in their community. $Individual_i$ measures the individual's support for the Taliban, where one suggests strong support and four indicates strong opposition. $Village_i$ indicates the wave-specific average of village support for the Taliban, excluding the respondent themselves. $Individual_i \times Village_i$ captures the marginal effect of joint variation in support for the Taliban. D_i indicates district-level fixed effects, W_t indicates wave-specific fixed effects, and X_i is a vector of control variables. Robust standard errors are clustered by district to account for the potential clustering of Taliban views within administrative units as well as reporting behavior. All models are adjusted using population sampling weights.

Table 1: Summary statistics

Variable	Mean	Std. dev.	Min.	Max.	*N*
Reported Taliban	0.588	0.492	0	1	9,850
Individual view of Taliban	3.417	0.814	1	4	9,850
Village view of Taliban	3.359	0.68	1	4	9,845
Pashtun	0.567	0.495	0	1	9,850
Shia	0.075	0.263	0	1	9,850
Age	35.114	11.749	18	85	9,850
Female	0.338	0.473	0	1	9,850
Married	0.861	0.346	0	1	9,850
Years of education	3.621	5.177	0	18	9,850
Employed	0.546	0.498	0	1	9,850
Student	0.035	0.184	0	1	9,850
Monthly income	3.253	1.142	1	7	9,850
Elder	0.074	0.261	0	1	9,850
Victim	0.308	0.462	0	1	9,803
Violence in village	0.474	0.499	0	1	9,794
Intimidation in village	0.458	0.498	0	1	9,806
Taxation in village	0.375	0.484	0	1	9,728
View of national army	1.68	0.794	1	4	9,809
View of national police	2.003	0.879	1	4	9,804
View of local police	2.222	0.934	1	4	9,738

Results

Of those individuals who had been exposed to illegal Taliban behavior, 58.8 percent indicated that they reported the crimes to internal or external authorities, which is consistent with other studies of reporting wrongdoing in a variety of contexts (e.g., Miceli et al. 1999; Miethe 1999). Approximately 58.7 percent of those in our sample report strong negative views of the Taliban, whereas 12.9 percent indicate that they support the Taliban. The remaining 28.4 percent hold weakly negative views of the Taliban. Similarly, when we aggregate community preferences, we find that roughly 14.8 percent of respondents' village views are at or below 2.5, which indicates indifference or support for the Taliban. Approximately 57 percent of village views are 3.5 or higher, indicating mixed or strong opposition to the Taliban. Although a disproportionate number of respondents reported negative views toward the Taliban, our sample size is large enough such that there are at least 700 respondents in each quadrant as depicted in Figure 1. All descriptive statistics for the variables used in the analysis can be seen in Table 1.

Looking at the control variables, more than half of our sample is Pashtun, which is the largest ethnic group in Afghanistan. Shia Muslims constitute 7.5 percent of the sample, with the remainder Sunni. Beyond those variables, the typical witness is 35 years old, male, and employed and has 3.6 years of schooling.[9] Elders constitute 7.4 percent of the sample. As elders constitute one of the internal authorities to whom individuals can report, it is unclear whether or not it is appropriate to include elders

within our sample. Although the subsequent analysis includes and controls for elders, all results are identical if elders are dropped.

Thirty-one percent of witnesses had been personally victimized by the Taliban at some point, largely due to attacks, recruitment, occupation, and looting. The next three variables indicate the type of crimes respondents witnessed in the previous month in their village. They come from the following questions: "How frequently have you experienced or seen violence in the last month in your community?" "How frequently have you experienced or seen intimidation in the last month in your community?" "In the last month, how frequently has the Taliban imposed zakat or taxation in your community?"[10] For each of these three questions, respondents indicated that the activity occurred "never," "once," "at least once a week," or "at least once a day." We collapse these options into binary variables that take a value of one if the particular activity occurred at least once in the previous month. Between 38 percent and 47 percent of respondents reported having observed each of these.

Finally, the last three rows of Table 1 indicate respondents' views of external social control agents. Similar to the question about the Taliban, respondents indicated on a scale of one to four whether they held very favorable, somewhat favorable, somewhat unfavorable, or very unfavorable views of the national army, the national police, and the local police. Overall, respondents had mixed opinions of governmental social control agents, although these lean in a slightly positive direction.

We next turn our attention to the regression-based evidence, based on the model specification in Equation (1). These results are presented in Table 2. We start with the simplest specification in column (1), with only district and wave fixed effects. Recall that *individual view* and *village view* range from one to four, with one representing strong support for the Taliban and four representing strong opposition to the Taliban. To account for the effect of social alignment, we interact these two parameters. Each base term captures the correlation between the parameter and the outcome when the other base term is at its lowest level (relative to the regression intercept). To ease interpretation, if our theoretical expectations are met, we would expect the base terms to be negative while the interaction is positive. If the base terms are negative, this suggests that the probability of reporting decreases as the individual's view becomes more misaligned with the village's view. If the interaction term is positive, this suggests that as the two parameters covary positively, the willingness to report increases. This is the pattern we observe in column (1). We visualize the marginal effects associated with this model specification in substantive terms in Figure 2(a). Notice that in the lower-left quadrant, where individual and village preferences for the Taliban are closely aligned, the probability of reporting nears 0.7. In the upper-right quadrant, where individual and village preferences against the Taliban are closely aligned, reporting approaches 0.6. In the socially misaligned quadrants (the upper left and lower right), the willingness to report drops below 0.5 and reaches a minimum of 0.44.

We next sequentially saturate the model specification with additional parameters including demographic controls (column (2)), measures of occupation (column (3)), past victimization (column (4)), type of exposure (column (5)), and attitudes toward Afghan security forces (column (6)). In line with our discussion above,

Table 2: Association between individual and village views of the Taliban and whether individual reports wrongdoing

	(1)	(2)	(3)	(4)	(5)	(6)
Individual view	−0.122[†]	−0.122[†]	−0.122[†]	−0.124[†]	−0.118[†]	−0.122[†]
	(0.041)	(0.041)	(0.041)	(0.041)	(0.040)	(0.040)
Village view	−0.127[†]	−0.127[†]	−0.126[†]	−0.128[†]	−0.122[†]	−0.130[†]
	(0.043)	(0.043)	(0.042)	(0.043)	(0.041)	(0.042)
Individual view × Village view	0.043[†]	0.043[†]	0.043[†]	0.044[†]	0.041[†]	0.043[†]
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Pashtun		0.037	0.038	0.044	0.041	0.048
		(0.034)	(0.033)	(0.033)	(0.034)	(0.032)
Shia		0.004	0.006	0.005	−0.000	0.000
		(0.050)	(0.049)	(0.050)	(0.048)	(0.047)
Age		−0.020[†]	−0.025[†]	−0.025[†]	−0.023[†]	−0.024[†]
		(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
Female		−0.019	−0.014	−0.016	−0.016	−0.012
		(0.025)	(0.033)	(0.033)	(0.033)	(0.033)
Married		0.026	0.022	0.023	0.021	0.023
		(0.021)	(0.022)	(0.022)	(0.022)	(0.022)
Years of education		0.005	0.003	0.002	0.002	0.001
		(0.009)	(0.010)	(0.010)	(0.009)	(0.010)
Employed			0.002	0.001	−0.001	−0.002
			(0.026)	(0.026)	(0.026)	(0.026)
Student			−0.019	−0.019	−0.013	−0.010
			(0.045)	(0.046)	(0.048)	(0.048)
Monthly income			0.009	0.009	0.007	0.007
			(0.008)	(0.008)	(0.008)	(0.008)
Elder			0.054	0.050	0.051	0.053
			(0.029)	(0.029)	(0.030)	(0.030)
Victim				0.021	0.031	0.031
				(0.023)	(0.023)	(0.022)
Violence in village					0.004	0.004
					(0.023)	(0.023)
Intimidation in village					0.013	0.015
					(0.019)	(0.019)
Taxation in village					−0.066[†]	−0.068[†]
					(0.020)	(0.020)
View of national army						−0.008
						(0.012)
View of national police						0.006
						(0.009)
View of local police						−0.017
						(0.009)
Parameters						
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Model statistics						
N	9,827	9,827	9,827	9,780	9,615	9,442

Notes: Outcome of interest is whether an individual reported wrongdoing. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. † $p < 0.01$; ∗ $p < 0.05$.

we incorporate these model parameters to account for potential concerns about systematic imbalances in the willingness to report crimes and individual preferences for the Taliban across various demographic groups as well as exposure to different types of Taliban activity, and views of Afghanistan army and police forces. In each of these cases, these omitted factors could potentially bias our findings. However, rather than attenuating our results, accounting for these potential sources of bias actually increases the magnitude of our estimates slightly. An examination of Figure 2, panels (b) through (f), reveals the robust substantive patterns present in the baseline specification, with social alignment leading to higher levels of reporting while social misalignment is associated with lower levels of reporting.

Figure 3 presents an alternative way of visualizing the results using model (6) of Table 2, which includes the full set of control variables. Each of the four lines indicates a different level of village attitudes toward the Taliban, the x axis indicates individual attitudes toward the Taliban, and the y axis indicates the predicted probability of reporting. Focusing first on the line where village views were equal to one (i.e., very favorable toward the Taliban), individuals who were in alignment with their village reported 70 percent of the time, whereas individuals whose views were in maximal misalignment with their village (i.e., very unfavorable toward the Taliban) reported 46 percent of the time. The pattern reverses, however, when village views were equal to four (i.e., very unfavorable toward the Taliban). In this case, 44 percent of individuals whose views were in maximal misalignment with their village (i.e., very favorable toward the Taliban) reported, whereas 59 percent of individuals whose views were in alignment with their village reported. Together, these effect sizes are substantive; individual and village views of the Taliban had a large effect on whether or not illegal Taliban behavior was reported.

The second component of our theoretical framework structures our next investigation: to whom do individuals report Taliban wrongdoing? From a cursory perspective, it is odd that individuals socially aligned with their communities in support of the Taliban would be *more* likely to report wrongdoing, not less. As we stress in our theoretical argument, however, alignment increases the willingness to report, but the authorities with whom individuals engage are likely to differ. When there is alignment in opposing the Taliban, we anticipate that reporting will be primarily channeled toward external authorities (the Afghan National Army or National Police), whereas individuals aligned in support of the Taliban are likely to engage with internal authorities (local village elders). We take advantage of an additional question in our survey that captures, conditional on having reported wrongdoing, to whom the subject reported.

After asking whether individuals reported illegal activity, the following question asks "To whom did you report it?" This was an open-ended question in which response options were not read to the respondents. Responses were then classified by the enumerator into nine different categories: "ANP" (Afghan National Police), "ANA" (Afghan National Army), "Elders," "Local warlord," "Local militia," "International forces (Foreign troops, Americans, etc.)," "Local government officials," "Ullamah" (local religious leaders), and "Taliban or other groups local commander/leader." Forty-three percent of respondents indicated that they reported to the Afghan National Police, and an additional 11 percent reported to the Afghan

(a) Benchmark specification

(b) + demographics

(c) + occupation

(d) + victimization

(e) + type of crime

(f) + view of ASF

Figure 2: Impact of individual and village views of the Taliban on willingness to report wrongdoing. *Notes:* Figure displays fitted values obtained from main regression specification (with additional parameters as noted). For interpretation, we focus on the corner of each quadrant. Bottom left (1,1) indicates internal alignment; upper right (4,4) indicates external alignment; off-diagonal corners (upper left, bottom right) indicate misalignment. Models are cumulative based on alphabetical sequence. ASF indicates Afghan Security Forces, which includes separate parameters for respondent attitudes toward the Afghan National Army, the Afghan National Police, and the local police.
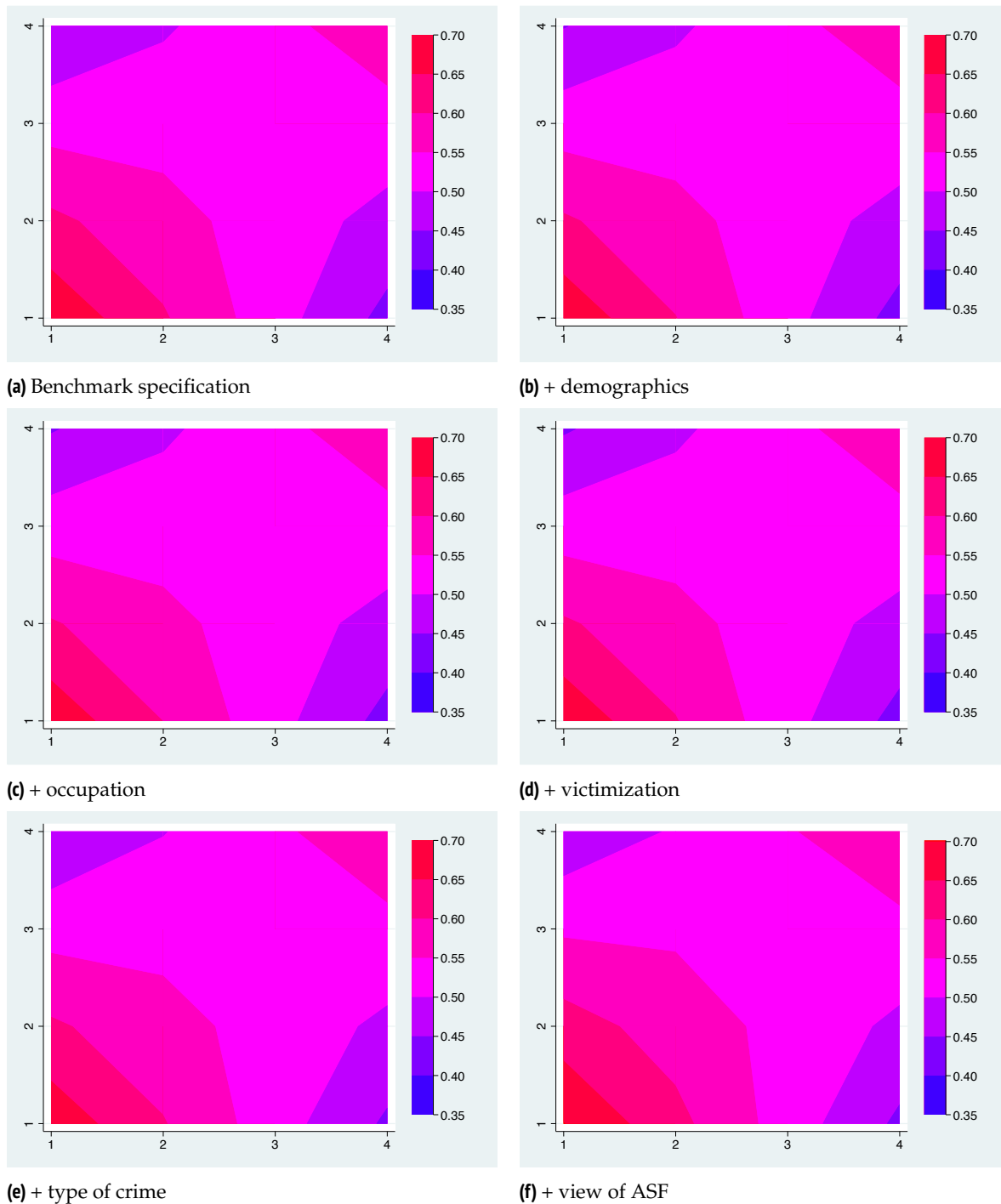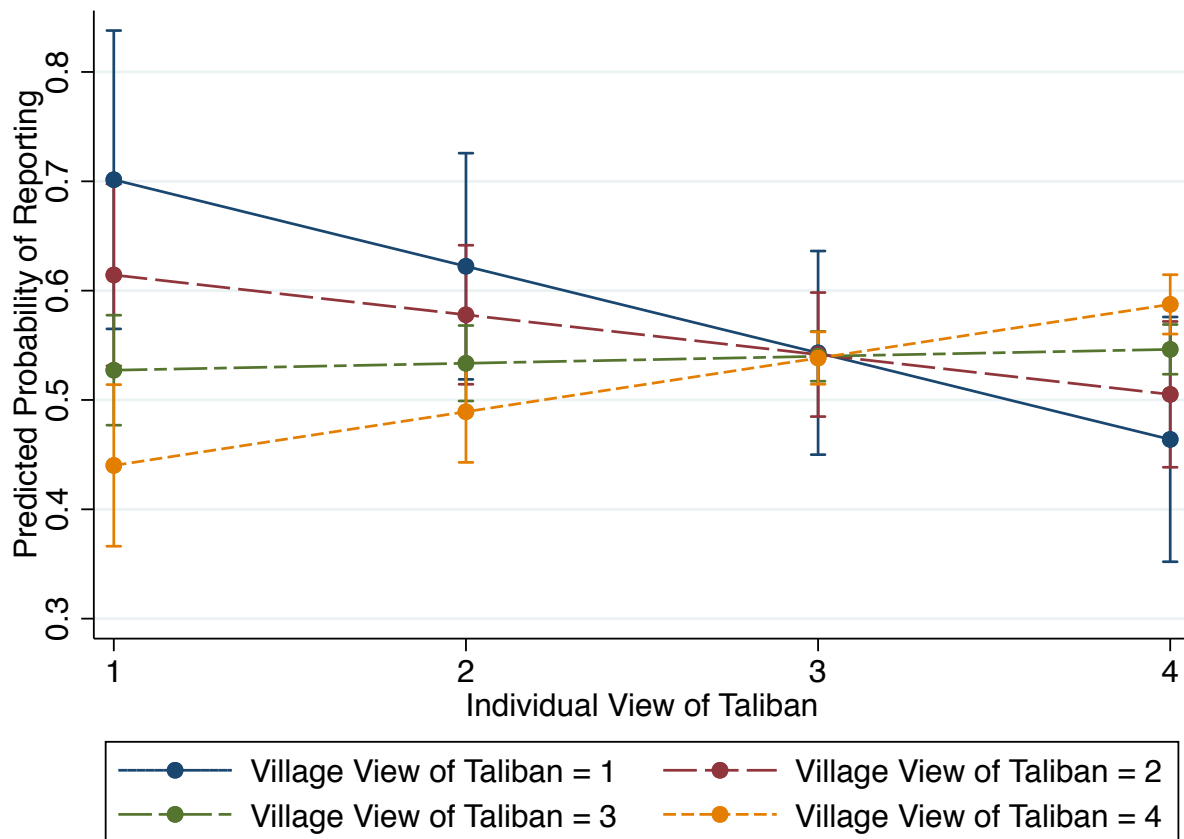
Figure 3: Predicted probability of individual and village views of the Taliban on willingness to report wrong-doing. *Notes:* Figure displays fitted values obtained from model (6) of Table 2 with 95 percent confidence intervals.

National Army. Thirty-four percent of respondents reported to village elders. Together this means that 88 percent of reporting was made to our key authorities of interest. Very few reports were made directly to the Taliban (0.5 percent), which is consistent with our interpretation of the reporting decision as being one of internal versus external reporting, not deciding between two rival forces. The next largest categories of reporting are local government officials (five percent) and Ullamah (three percent). Only 0.1 percent of respondents reported to international forces.

Based on our theoretical interest in distinguishing between internal and external reporting, we specify reporting outside the village as occurring when individuals report to the Afghan National Police or National Army, and reporting inside the village as when individuals report to village elders. As mentioned above, this constitutes 88 percent of all reports. Our results, however, are not sensitive to the inclusion of the other categories (e.g., including international forces and local government officials in external reporting and Ullamah in internal reporting).

We begin with a descriptive examination of the data on reporting wrongdoing to internal or external authorities in each of the four quadrants in Figure 1. Consistent with our theory, we are primarily interested in patterns of reporting in the bottom-left and upper-right cells. This allows us to focus squarely on the shift from internal to external alignment. Conditional on reporting, those having negative views of the Taliban (i.e., values of three or four) and living in aligned villages (i.e., having values of three or more) reported externally 57 percent of the time, whereas those having positive views of the Taliban (i.e., values of one or two) and living in aligned villages (i.e., having values of two or less) reported externally only 40 percent of the time. If we look at more extreme levels of alignment (i.e., individuals and villages that are either both strongly negative, with values of 3.5 or greater, or both strongly positive, with values of 1.5 or less), those percentages become 61 percent and 31 percent, respectively. In other words, individuals in villages aligned in strong opposition to the Taliban were almost twice as likely to report to external authorities as those in villages aligned in strong favor of the Taliban.

In order to verify this pattern, we assess whether a regression-based approach provides findings consistent with our descriptive evidence above. Mirroring the descriptive approach, we condition our analysis on reporting and those who fall within the two aligned quadrants in Figure 1. We introduce these results in Table 3. This table replicates the benchmark specification and robustness checks presented in Table 2. Notice, in line with our expectations, that reporting to external authorities is increasing with aligned opposition to the Taliban. In the first specification, without accounting for potential bias from individual-level characteristics, our estimate is not precise at the five percent level (although the p value is 0.06). As we add controls and account for potentially confounding factors, the estimated effect of interest (*individual view $\times$ village view*) becomes and remains statistically precise. The interaction effect confirms that as we move from the bottom-left corner of Figure 1 to the top-right corner, individuals become increasingly likely to report to external authorities.

Taken together, the results from Table 2 and Table 3 yield a compelling pattern. Individuals are most likely to report wrongdoing when their views of the Taliban align with other members of their community. Those that are aligned in opposition to the Taliban are more likely to report to external authorities, whereas individuals aligned in favor of the Taliban are more likely to report to internal authorities.

These results also confirm that deciding to whom to report is not simply a matter of which authority is likely to be most successful at punishing and deterring the offending behavior. District fixed effects account for the variation in Taliban presence and power throughout the country, such that our results should not be driven by varying abilities of different authorities in different areas to resolve problems. Furthermore, if effectiveness is the only concern of observers of wrongdoing, then misaligned individuals should not hesitate to report to whichever authority they perceive as the most effective source of social control. The fact that these individuals are less likely to report suggests that it is community dynamics—and not perceived effectiveness—that largely motivates which reporting channel is activated.

Table 3: Individual and village views of the Taliban and to whether individual reports wrongdoing to external authorities (compared to internal authorities)

	(1)	(2)	(3)	(4)	(5)	(6)
Individual view	−0.167	−0.178	−0.183	−0.186	−0.178	−0.173
	(0.116)	(0.113)	(0.113)	(0.112)	(0.100)	(0.095)
Village view	−0.139	−0.148	−0.153	−0.154	−0.163	−0.214*
	(0.108)	(0.105)	(0.105)	(0.104)	(0.094)	(0.101)
Individual view × Village view	0.071	0.074*	0.075*	0.076*	0.076*	0.080*
	(0.038)	(0.037)	(0.037)	(0.037)	(0.032)	(0.033)
Pashtun		0.067	0.069	0.070	0.057	0.062
		(0.038)	(0.037)	(0.037)	(0.037)	(0.038)
Shia		0.133*	0.129*	0.129*	0.127*	0.119*
		(0.060)	(0.059)	(0.060)	(0.060)	(0.059)
Age		−0.003	0.003	0.003	−0.000	−0.001
		(0.011)	(0.012)	(0.011)	(0.011)	(0.011)
Female		−0.087†	−0.086*	−0.089*	−0.079*	−0.089*
		(0.026)	(0.040)	(0.041)	(0.040)	(0.039)
Married		0.024	0.032	0.030	0.020	0.023
		(0.035)	(0.033)	(0.033)	(0.032)	(0.032)
Years of education		0.004	0.002	0.002	−0.000	−0.007
		(0.011)	(0.012)	(0.012)	(0.013)	(0.013)
Employed			0.007	0.004	0.015	0.011
			(0.034)	(0.034)	(0.034)	(0.034)
Student			0.051	0.043	0.057	0.045
			(0.063)	(0.063)	(0.059)	(0.059)
Monthly income			0.013	0.013	0.015	0.007
			(0.010)	(0.010)	(0.010)	(0.010)
Elder			−0.081	−0.085*	−0.078	−0.098*
			(0.041)	(0.041)	(0.043)	(0.046)
Victim				0.017	−0.019	−0.023
				(0.027)	(0.027)	(0.028)
Violence in village					−0.015	−0.015
					(0.039)	(0.038)
Intimidation in village					0.099*	0.103*
					(0.040)	(0.039)
Taxation in village					0.096†	0.080†
					(0.026)	(0.026)
View of national army						−0.074†
						(0.016)
View of national police						−0.025
						(0.015)
View of local police						−0.013
						(0.012)
Parameters						
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Model statistics						
N	4,411	4,411	4,411	4,399	4,350	4,305

Notes: Outcome of interest is whether an individual reported wrongdoing to external (as opposed to internal) authorities, conditional on reporting and in-group and out-group quadrants of Taliban support. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. † $p < 0.01$; ∗ $p < 0.05$.

Supplemental Tests

In this section, we consider a number of supplemental tests. We assess (1) balance in perceptions of crime, (2) exposure to different types of crime, and (3) measurement and construct validity; (4) conduct randomization inference tests; and assess (5) robustness to incorporating measures of village structure as well as (6) convergence between village preferences and neighboring areas.

Balance Tests

We begin by investigating an important concern about how attitudes toward the Taliban may influence perceptions of illegal behavior. It is possible that support for or opposition to the Taliban affects individuals' interpretations of particular behaviors as illegal or not. In other words, those opposed to the Taliban may more readily interpret certain Taliban behaviors as illegal, whereas those who support the Taliban may not. If this is the case, then our data are subject to selection bias regarding who observes illegal behavior. We therefore reevaluate our results from Table 2 using a two-stage estimation procedure developed by Heckman (1979) to account for this issue. The first stage uses a probit model that describes the propensity of having observed illegal behavior among all survey respondents controlling for the full set of independent and control variables, excepting those that may be determined coterminously with the outcome (e.g., the type of illegal behavior and whether the observer is a victim). The second stage includes the inverse Mills ratio of the first-stage coefficients in order to mitigate selection bias. Table 4 displays these results. The coefficients of interest—individual and group attitudes toward the Taliban, along with their interaction—are all similar in magnitude and statistical significance to the results in Table 2. The resultant predicted probabilities are also almost identical to the values displayed in Figure 3.

Another natural concern is whether subjects and villages with different levels of Taliban support are exposed to different *types* of crimes, not just to crimes in general. For example, perhaps areas with high Taliban support are subject to taxation, whereas areas with low Taliban support are subject to violence. If that is the case, then differential exposure to particular crimes could be biasing our results. Although we control for the type of crime in model (5) of Table 2 and Table 3, here we give this issue more focused attention. To examine these potential imbalances, we repeat our benchmark model specification and evaluate whether subjects with different views of the Taliban are exposed to different illegal activities.

We present these results in Table 5. We show the types of crimes—intimidation, violence, and illegal taxation—listwise in columns (1) through (3). Our evidence consistently suggests that there is no statistically significant imbalance across villages (and individuals) with exposure to Taliban crimes across our core measures of how individuals and villages view the Taliban. This suggests that those with differing views toward the Taliban were not exposed to different crimes, lending support to our interpretation of the results.

Table 4: Association between individual and village views of the Taliban and whether individual reports wrongdoing (two-stage estimation)

	(1)	(2)	(3)	(4)	(5)	(6)
Individual view	−0.122[†]	−0.120[†]	−0.108[†]	−0.111[†]	−0.105*	−0.121[†]
	(0.042)	(0.041)	(0.041)	(0.042)	(0.041)	(0.041)
Village view	−0.126[†]	−0.131[†]	−0.161[†]	−0.164[†]	−0.158[†]	−0.132*
	(0.043)	(0.045)	(0.052)	(0.052)	(0.049)	(0.055)
Individual view × Village view	0.042[†]	0.043[†]	0.042[†]	0.043[†]	0.041[†]	0.043[†]
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Pashtun		0.044	0.079	0.085	0.082	0.050
		(0.034)	(0.044)	(0.044)	(0.043)	(0.053)
Shia		0.002	−0.013	−0.013	−0.018	−0.001
		(0.049)	(0.052)	(0.052)	(0.050)	(0.052)
Age		−0.020[†]	−0.018	−0.018	−0.016	−0.023
		(0.008)	(0.011)	(0.011)	(0.011)	(0.012)
Female		−0.015	−0.018	−0.019	−0.020	−0.013
		(0.028)	(0.033)	(0.033)	(0.034)	(0.033)
Married		0.026	−0.015	−0.014	−0.015	0.021
		(0.023)	(0.038)	(0.038)	(0.038)	(0.048)
Years of education		0.006	0.013	0.011	0.011	0.001
		(0.010)	(0.012)	(0.012)	(0.011)	(0.012)
Employed			0.027	0.025	0.024	−0.000
			(0.035)	(0.035)	(0.035)	(0.041)
Student			−0.076	−0.076	−0.069	−0.013
			(0.063)	(0.063)	(0.062)	(0.074)
Monthly income			0.022	0.022	0.020	0.008
			(0.012)	(0.012)	(0.012)	(0.016)
Elder			0.025	0.021	0.021	0.051
			(0.039)	(0.039)	(0.039)	(0.047)
Victim				0.023	0.032	0.031
				(0.023)	(0.022)	(0.022)
Violence in village					0.003	0.004
					(0.023)	(0.023)
Intimidation in village					0.016	0.015
					(0.019)	(0.019)
Taxation in village					−0.069[†]	−0.068[†]
					(0.020)	(0.020)
View of national army						−0.008
						(0.012)
View of national police						0.006
						(0.010)
View of local police						−0.017
						(0.011)
Parameters						
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Model statistics						
N	9,641	9,641	9,641	9,597	9,442	9,442

Notes: Outcome of interest is whether an individual reported wrongdoing. The inverse Mills ratio is included in all models and comes from the estimated coefficients of the first-stage model to correct for the sampling bias of being exposed to illegal behavior. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. † $p < 0.01$; * $p < 0.05$.

Table 5: Assessing balance in the probability of exposure to type of wrongdoing with respect to individual and village views of the Taliban

	(1) Intimidation in village	(2) Violence in village	(3) Taxation in village
Individual view	−0.086	−0.104	0.096
	(0.070)	(0.064)	(0.066)
Village view	−0.051	−0.061	0.066
	(0.073)	(0.065)	(0.071)
Individual view × Village view	0.023	0.028	−0.030
	(0.023)	(0.020)	(0.021)
Pashtun	−0.026	−0.049	0.017
	(0.027)	(0.028)	(0.033)
Shia	0.019	−0.005	−0.087
	(0.045)	(0.046)	(0.054)
Age	0.007	0.006	0.010
	(0.006)	(0.007)	(0.008)
Female	−0.019	−0.034	−0.060*
	(0.023)	(0.024)	(0.023)
Married	0.009	−0.004	0.003
	(0.017)	(0.017)	(0.017)
Years of education	0.013	0.016	−0.004
	(0.008)	(0.008)	(0.013)
Employed	0.005	−0.016	−0.043
	(0.019)	(0.019)	(0.024)
Student	−0.008	−0.100†	−0.006
	(0.040)	(0.036)	(0.040)
Monthly income	0.002	−0.005	−0.007
	(0.008)	(0.008)	(0.010)
Elder	−0.031	−0.003	0.016
	(0.029)	(0.028)	(0.025)
View of national army	0.013	0.014	−0.025*
	(0.012)	(0.011)	(0.012)
View of national police	0.021*	0.011	−0.013
	(0.010)	(0.010)	(0.012)
View of local police	−0.017*	−0.005	0.022*
	(0.008)	(0.008)	(0.009)
Parameters			
District fixed effects	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes
Model statistics			
N	9,602	9,591	9,528

Notes: Outcome of interest is whether an individual is exposed to wrongdoing (with specific type noted in text). Heteroskedasticity robust standard errors clustered by district are reported in parentheses. † $p < 0.01$; ∗ $p < 0.05$.

Measurement and Construct Validity

Survey data can be unreliable due to reporting bias: subjects may respond to direct questions about controversial topics in a manner that veils their true preferences. Although the survey does not use list or endorsement methods (Blair, Imai, and Lyall 2014), we are able to corroborate the direct attitude responses with indirect measures, which provide some confidence that responses are accurate.

To investigate this issue, we evaluate whether a direct survey measure is unreliable by studying whether responses across direct and indirect questions correspond to one another. With indirect questions, subjects may be more willing to reveal their true preferences. Fortunately, our data give us several opportunities to conduct this analysis. In addition to the direct measure of Taliban support, survey subjects were also asked about their views of people who support the Taliban. More specifically, they were asked (separately) to what extent they agreed or disagreed that "People who support the Taliban are Afghan patriots," "People who support the Taliban want peace for all of Afghanistan," and "People who support the Taliban do not feel that the Afghan Government cares about their needs." Similar to other questions, these statements were rated as "Strongly Agree," "Agree," "Disagree," and "Strongly Disagree," with higher values meaning increasing disagreement with the statement. Notice that individuals were not asked directly about their own preferences for the Taliban in these questions. Rather, the questions are framed broadly as indirect assessments of their views of others who support the Taliban.

To test whether the direct and indirect measures map onto one another, we use the most demanding specification presented in Table 2. These results are reported in Table 6. We are specifically interested in the correlation reported for the subject's own support for the Taliban. The outcome variable changes by column and is noted in each heading. For each model, increasingly negative direct views of the Taliban are associated with increasingly negative indirect views of the Taliban. These results provide evidence of a correspondence between a respondent's own support for the Taliban and how they view others with similar viewpoints. This suggests that respondents were not suffering from social desirability bias. Even if responses were biased, this would likely inflate overall reporting, perhaps particularly to external authorities. However, we still would not expect to see the hypothesized interaction effect, nor the distinction between internal and external reporting as a function of individual and group attitudes.

Randomization Inference

It is also possible that our approach to drawing inferences about the precision of our estimates is imperfect. In particular, we calculate cluster robust standard errors that account for heteroskedasticity within administrative districts. This is a reasonable approach because the administrative district is the primary sampling unit from which villages and settlements are selected in the second phase of randomization. If individual preferences, community attitudes, and Taliban presence are spatially correlated within district boundaries, this approach helps adjust our inferences about statistical precision (standard errors) accordingly. An alternative approach is to use randomization inference, which does not require either a strict null hypothesis of

Table 6: Association between main measure of Taliban support and individual's indirect assessment of Taliban supporters

| | (1) | (2) | (3) |
| | | | Government does |
	Patriots	Want peace	not care
Individual view	0.283†	0.255†	0.138†
	(0.026)	(0.023)	(0.031)
Pashtun	−0.018	−0.035	−0.005
	(0.063)	(0.053)	(0.047)
Shia	0.073	−0.003	−0.104
	(0.096)	(0.088)	(0.125)
Age	0.013	0.004	0.003
	(0.014)	(0.014)	(0.016)
Female	0.044	0.019	0.015
	(0.087)	(0.081)	(0.056)
Married	0.037	0.053	0.040
	(0.044)	(0.044)	(0.037)
Years of education	0.008	−0.021	0.008
	(0.018)	(0.017)	(0.018)
Employed	−0.006	−0.039	0.011
	(0.061)	(0.057)	(0.048)
Student	0.095	0.133	−0.119
	(0.085)	(0.092)	(0.085)
Monthly income	0.057†	0.029	−0.019
	(0.018)	(0.017)	(0.020)
Elder	−0.114*	−0.054	−0.028
	(0.056)	(0.050)	(0.052)
Victim	0.002	−0.068†	−0.007
	(0.032)	(0.025)	(0.032)
Violence village	0.104*	0.079	0.080
	(0.041)	(0.044)	(0.044)
Intimidation in village	0.038	0.077	0.090
	(0.041)	(0.051)	(0.047)
Taxation in village	−0.025	0.027	−0.021
	(0.035)	(0.037)	(0.032)
View of national army	0.021	−0.013	0.044
	(0.023)	(0.022)	(0.023)
View of national police	−0.001	0.037	−0.014
	(0.021)	(0.020)	(0.018)
View of local police	−0.010	−0.025	0.010
	(0.014)	(0.018)	(0.022)
Parameters			
District fixed effects	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes
Model statistics			
N	9,410	9,404	9,326

Notes: Outcome of interest is individual's indirect assessment of Taliban supporters. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. † $p < 0.01$; ∗ $p < 0.05$.

(a) Individual view   **(b)** Village view   **(c)** Individual view $\times$ Village view
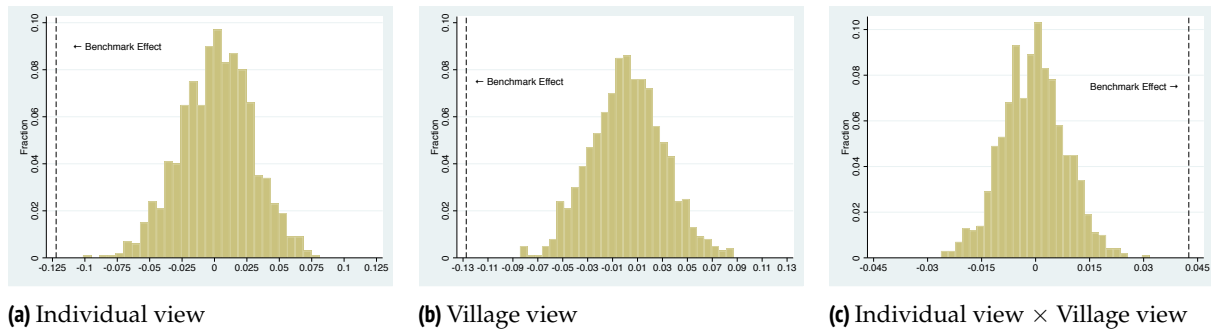
Figure 4: Randomization inference to evaluate robustness of benchmark effect. *Notes:* We randomly reshuffle the outcome vector (reporting wrongdoing) in our data using a spatial autocorrelation adjusted randomization inference test ($\times$ 1,000). In (a), we plot the corresponding output for the base term *individual view*. In (b), we plot the corresponding output for the base term *village view*. In (c), we plot the corresponding output for the base term *individual view* $\times$ *village view*. The model specification is equivalent to Table 2, column (4).

zero or any strict assumptions about spatial correlation. However, given that we know Taliban activities were spatially concentrated in the southern (Helmand, Kandahar) and southeastern (Nangarhar) regions of the country, simple randomization inference will lead us to draw misleadingly confident inferences about precision. To account for this, we employ a spatial correlation corrected randomization inference technique. The intuition of the test is straightforward. For a given set of data, one can randomly permute the sequence of outcomes, assigning the willingness to report wrongdoing from one individual to another, and replicate the benchmark regression specification. If this exercise is done a sufficient number of times and the research design follows standard practice, the distribution of corresponding regression coefficients should be normally distributed around zero. Calculating the *p* value of a regression coefficient then involves calculating the percentage of random draws that produce coefficient estimates at least as extreme as the one observed in the original unshuffled data. To correct for spatial correlation, we stratify the randomization of individual responses by administrative district. This provides a secondary method for assessing the precision of our findings even in the presence of spatial dependence in individual and community preferences and provides additional confidence in the statistical relevance of any associations we identify.

We introduce these results in Figure 4, panels (a) through (c). Notice that in each case, the distribution of parameter estimates is centered around zero and approximately normally distributed. In each case, there are no observed values more extreme than the ones we estimate in our main specification. This is true even in the presence of spatial correction and suggests that our main estimates are highly precise and unlikely to occur by random chance given the large number of random shuffling of the data ($s = 1,000$).

Village Structure and Community Clusters

The spatial precision of our survey data enables us to conduct two more novel tests. The first uses information about the sequence of surveys administered within each village. As described above, enumerators randomly administered their first survey in each village at one of five points: the eastern border of the village, the western border, the northern border, the southern border, or the village center. This means that 80 percent of enumerator random walks started from the outskirts of the village and progressed toward the center, whereas the other 20 percent did the opposite. Additionally, qualitative evidence from communities across rural Afghanistan suggests that elders are most likely to reside close to the center of the village. We use these two pieces of information to determine when an enumerator began on the outskirts or center of a village, depending on whether an elder appears early in the sequence or towards the end (in those villages in which an elder was surveyed). When an elder was surveyed early on, it is more likely that this village belongs to the 20 percent of cases where the sequence began in the center of the village. By knowing the path that enumerators took, (inside-out vs. outside-in) we can generate a potentially useful measure of geographic centrality within a village. It is possible that this is an important omitted variable that mediates the formation of individual preferences vis-á-vis the social environment within which individuals are embedded. In other words, perhaps misaligned individuals are socially marginalized and live on the outskirts of their villages, and their low status (rather than misalignment) is what decreases their willingness to report.

We incorporate this measure in Table 7 as a control variable, ranging from one to five, where higher numbers indicate greater geographic centrality within a village. We find evidence that even after accounting for village structure, our main estimates are largely unchanged. Even though village structure influences the willingness to report wrongdoing, it does not systematically reduce the effects of our key independent variables.

Another robustness check of our results involves relying on spatial details regarding the location of sampled communities in order to construct a proximity-based measure of the wider social setting within which each village is situated. This allows us to capture potential clusters of communities that exhibit congruent environments. For nearly all of the sampled sites, we have exact latitude and longitude coordinates. Using geographic mapping software, we use this information to identify the cluster of 10 villages most proximate to a given sampling site in a given wave. Based on these spatial linkages, we can then construct aggregate measures of Taliban views across the social cluster of villages centered around each given sampling point. If, for example, the administrative district is an insufficient unit for capturing the spatial correlation in Taliban views around any given individual and their village, this alternative approach is a much more natural social object: the community of communities within which a person and their neighbors are embedded.

We visualize a theoretical approach to these village clusters in Figure 5. The physical pattern of clustering depends on the geography within which a given sampled settlement is located vis-á-vis neighboring villages. For each village in a given wave, the corresponding cluster is calculated conditional on the sampled

Table 7: Association between individual and village views of the Taliban and whether individual reports wrongdoing, accounting for village structure

	(1)	(2)	(3)	(4)	(5)	(6)
Individual view	$-0.123^{\dagger}$	$-0.123^{\dagger}$	$-0.122^{\dagger}$	$-0.125^{\dagger}$	$-0.118^{\dagger}$	$-0.123^{\dagger}$
	(0.041)	(0.041)	(0.041)	(0.041)	(0.040)	(0.040)
Village view	$-0.128^{\dagger}$	$-0.128^{\dagger}$	$-0.127^{\dagger}$	$-0.129^{\dagger}$	$-0.123^{\dagger}$	$-0.131^{\dagger}$
	(0.043)	(0.043)	(0.042)	(0.043)	(0.041)	(0.042)
Individual view × Village view	$0.043^{\dagger}$	$0.043^{\dagger}$	$0.043^{\dagger}$	$0.044^{\dagger}$	$0.041^{\dagger}$	$0.043^{\dagger}$
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Village centrality	$0.012^{\dagger}$	$0.013^{\dagger}$	$0.011^{\dagger}$	$0.011^{\dagger}$	$0.011^{\dagger}$	$0.010^{\dagger}$
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Pashtun		0.040	0.040	0.046	0.043	0.050
		(0.033)	(0.033)	(0.033)	(0.033)	(0.032)
Shia		0.006	0.007	0.007	0.001	0.001
		(0.050)	(0.050)	(0.050)	(0.048)	(0.047)
Age		$-0.022^{\dagger}$	$-0.026^{\dagger}$	$-0.025^{\dagger}$	$-0.024^{\dagger}$	$-0.024^{\dagger}$
		(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
Female		-0.018	-0.014	-0.015	-0.015	-0.012
		(0.025)	(0.033)	(0.033)	(0.033)	(0.033)
Married		0.024	0.021	0.021	0.020	0.022
		(0.021)	(0.022)	(0.022)	(0.022)	(0.021)
Years of education		0.005	0.003	0.002	0.002	0.001
		(0.009)	(0.010)	(0.010)	(0.009)	(0.010)
Employed			0.003	0.002	-0.000	-0.001
			(0.025)	(0.025)	(0.026)	(0.026)
Student			-0.017	-0.018	-0.012	-0.008
			(0.045)	(0.045)	(0.048)	(0.047)
Monthly income			0.009	0.009	0.007	0.007
			(0.008)	(0.008)	(0.008)	(0.008)
Elder			0.040	0.037	0.038	0.040
			(0.030)	(0.029)	(0.030)	(0.030)
Victim				0.021	0.030	0.031
				(0.023)	(0.023)	(0.022)
Violence in village					0.004	0.004
					(0.023)	(0.023)
Intimidation in village					0.014	0.016
					(0.019)	(0.019)
Taxation in village					$-0.066^{\dagger}$	$-0.068^{\dagger}$
					(0.020)	(0.020)
View of national army						-0.008
						(0.012)
View of national police						0.006
						(0.009)
View of local police						-0.017
						(0.009)
Parameters						
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Model statistics						
N	9,827	9,827	9,827	9,780	9,615	9,442

Notes: Outcome of interest is whether an individual reported wrongdoing (conditional on being exposed). Heteroskedasticity robust standard errors clustered by district are reported in parentheses. $\dagger\ p < 0.01$; $*\ p < 0.05$.

(a) Cluster 1   **(b)** Cluster 2   **(c)** Cluster 3
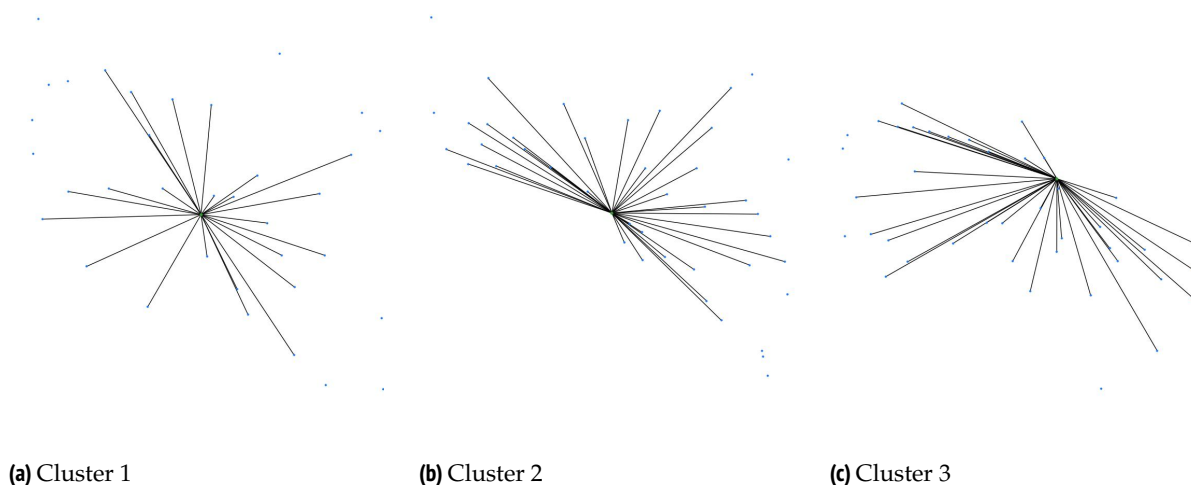
Figure 5: Using precise village locations to construct clusters. *Notes:* We illustrate the village clustering process above. Using locations of each sampled village and the sampled subset of the sampling frame (of villages), we construct clusters.

frame (the villages surveyed) that is randomized.[11] We measure these clusters, aggregate preferences across the entire cluster (holding out the sampled unit itself), and incorporate this measure of aggregated preferences as a control variable in the model specifications in Table 8. The central concern we are attempting to address here is the possibility that village-wide preferences simply reflect the larger social system within which individuals are embedded and thus our inferences about social alignment are confounded by the covariance between village and community cluster-level preferences. If this is a meaningful issue from an inferential perspective, our estimated effects are likely biased upward because our measure of village views is likely positively correlated with cluster-level measures. Our results, however, suggest that the spatial clustering in preferences is unlikely to cause bias as our estimated main effects are highly consistent and statistically indistinguishable from the core results presented above.

Overall, these supplemental tests give us additional confidence in the consistency of our core finding: social alignment influences when individuals are willing to report wrongdoing and to whom they report it. Additionally, the effect sizes remain consistent across all variations, providing added confidence in the results.

Discussion

In this article, we presented a theory explaining when individuals report wrongdoing to internal authorities and when they report to external authorities. When there is negative alignment between the witness and the community about the salient category of the wrongdoer, the witness is likely to report to external authorities. When there is positive alignment between the witness and the community about the salient category of the wrongdoer, the witness is likely to report to internal

Table 8: Association between individual and village views of the Taliban and whether individuals report wrongdoing, accounting for village clusters

	(1)	(2)	(3)	(4)	(5)	(6)
Individual view	−0.122†	−0.124†	−0.123†	−0.126†	−0.119†	−0.124†
	(0.041)	(0.041)	(0.041)	(0.041)	(0.039)	(0.039)
Village view	−0.122†	−0.124†	−0.123†	−0.125†	−0.117†	−0.123†
	(0.045)	(0.045)	(0.045)	(0.045)	(0.042)	(0.043)
Individual view × Village view	0.043†	0.044†	0.044†	0.045†	0.043†	0.044†
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Surrounding villages' view	0.043	0.045	0.047	0.046	0.047	0.050
	(0.041)	(0.041)	(0.041)	(0.041)	(0.039)	(0.039)
Pashtun		0.028	0.029	0.034	0.027	0.036
		(0.029)	(0.028)	(0.028)	(0.028)	(0.027)
Shia		−0.022	−0.018	−0.019	−0.023	−0.021
		(0.054)	(0.053)	(0.054)	(0.052)	(0.051)
Age		−0.018*	−0.024†	−0.024†	−0.022†	−0.023†
		(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
Female		−0.043	−0.048	−0.050	−0.051	−0.046
		(0.025)	(0.035)	(0.035)	(0.035)	(0.035)
Married		0.027	0.024	0.025	0.026	0.029
		(0.024)	(0.025)	(0.026)	(0.026)	(0.025)
Years of education		0.005	0.003	0.002	0.002	0.001
		(0.009)	(0.010)	(0.010)	(0.009)	(0.010)
Employed			−0.011	−0.013	−0.016	−0.015
			(0.027)	(0.027)	(0.027)	(0.027)
Student			−0.031	−0.031	−0.027	−0.020
			(0.046)	(0.046)	(0.047)	(0.046)
Monthly income			0.008	0.007	0.005	0.005
			(0.008)	(0.008)	(0.008)	(0.009)
Elder			0.065*	0.061*	0.061	0.061*
			(0.031)	(0.031)	(0.031)	(0.031)
Victim				0.019	0.033	0.032
				(0.025)	(0.024)	(0.024)
Violence in village					−0.009	−0.007
					(0.021)	(0.021)
Intimidation in village					0.012	0.013
					(0.020)	(0.020)
Taxation in village					−0.073†	−0.074†
					(0.020)	(0.020)
View of national army						−0.013
						(0.013)
View of national police						0.012
						(0.010)
View of local police						−0.020*
						(0.009)
Parameters						
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Model statistics						
N	8,702	8,702	8,702	8,665	8,526	8,374

Notes: Outcome of interest is whether an individual reported wrongdoing (conditional on being exposed). Heteroskedasticity robust standard errors clustered by district are reported in parentheses. † $p < 0.01$; * $p < 0.05$.

authorities. When misalignment occurs, the wrongdoing is less likely to be reported at all. Although questions of whether to report and where to report are typically treated separately, this work reveals that they are to some extent intertwined. The same factors that can push someone to report to different authorities can also lead to not reporting at all.

Two primary consequences emerge from these findings. The first is that even egregious crimes may be dealt with internally depending on attitudes toward the category affiliation of offenders, effectively shielding certain perpetrators from outside scrutiny and investigation (Alexander 2018; Palmer and Feldman 2018). This helps to explain how wrongdoing is often able to stay hidden from public view for such a long time, even when large numbers of organizational members are aware of it. Hundreds of priests knew about the prevalence of child abuse within the Catholic Church, and more than a hundred employees were involved in Enron's securities fraud (Aven 2015; Boston Globe 2003). Yet law enforcement did not become aware of either scandal for years because reporting was almost exclusively internal to the organizations. From corporate misconduct to sexual abuse to illegal behavior by the Taliban, wrongdoing can simultaneously be widely known and hidden, which suggests one of the reasons why wrongdoing is increasingly recognized as a normal part of organizational behavior (Palmer 2012).

Second, although some perpetrators of wrongdoing may be shielded by internal social control, others do not share the same opportunities. Basing the reporting decision on attitudes toward perpetrator categories may lead to members of minorities or disadvantaged categories being disproportionately reported externally, whereas members of majorities or advantaged categories are reported internally. Even in cases of perceived minor violations, transgressors from disadvantaged categories may be channeled to external authorities who may be more punitive. For example, in 2018 a White student reported a Black student to the police—rather than to campus officials—at Yale University for suspected trespassing because she was taking a nap in the common area of their dorm (Wootson 2018). Such reporting biases can lead to perpetrators of the same alleged crime having dramatically different experiences with institutions of criminal justice. Even a relatively fair justice system may produce biased outcomes so long as different categories of people are reported to different authorities.

One way to perhaps mitigate some of the aforementioned negative consequences is for internal and external authorities to work together and share information about allegations of illegal behavior. By doing so, egregious crimes will less easily be shielded internally, and inequality may be reduced regarding access to particular sources of social control. In fact, exactly this form of cooperation has been proposed in Afghanistan as a solution to the policing failures of the national government (Wardak 2011; Wardak et al. 2007). Considering that reporting behavior is embedded within group dynamics, cooperation between nested social control agents is likely to be more effective than encouraging witnesses to report to one authority or the other.[12]

Such a solution, however, does little about cases of misalignment, where individuals do not feel comfortable reporting to any authority at all. Unfortunately, misalignment might be particularly likely among those personally victimized by

wrongdoing. The very act of victimization may lead the victim to view the perpetrator and the perpetrator's category in a negative light while leaving the views of other group members unchanged. For example, individuals who are sexually assaulted by majority members at college or in the workplace may be hesitant to go to the police because they fear retaliation or ostracism from other group members. At the same time, victims may fear that they will not be believed or taken seriously enough if they report internally, especially if the wrongdoer has high status within the group. This may lead to victims of crimes—especially those lacking concrete evidence—to not report wrongdoing at all. Consistent with this, the proportion of those victimized by the Taliban is greatest in the upper-left quadrant of Figure 1 where individuals dislike the Taliban and are misaligned with their communities. Thirty-six percent of individuals in that quadrant were previously victimized, compared with less than 25 percent for each of the other three quadrants. Certainly, this topic deserves further study.

The theory is potentially applicable to non-formal means of social control. Informal means, such as norm enforcement, are a critical way by which groups regulate the behavior of their members. Norm enforcement involves sanctioning behavior that is kept private within the group and is likely to remain outside the awareness of both internal and external authorities. In this way, consistently sanctioning a particular category through norm enforcement likely represents an even stronger act of affiliation than reporting to internal authorities. Norms apply primarily to group members (Akerstrom 1988), so the use of formal reporting channels (rather than informal sanctioning) might be considered a symbolic act of disaffiliation. One can potentially conceive of three concentric circles of enforcement—the innermost representing informal norm enforcement, the middle representing internal reporting, and the outer representing external reporting—which are decreasing in the extent to which transgressors and transgressor categories are symbolically affiliated with the group. From this perspective, norm enforcement becomes a means of defining group boundaries, not just by emphasizing group norms, but by indicating who should be subject to them.

Several scope conditions limit the theory. For one, the witnesses must value their membership in the group. Only in this case do the views of other members become particularly salient. For example, an individual's willingness to inform external authorities about wrongdoing is likely much greater if that individual is no longer a member of the group. Second, the theory assumes that there is consensus about what the salient category or categories of perpetrators are. Given the primacy that categories such as race and gender typically have, we expect this assumption to often be reasonable (Van Knippenberg and Dijksterhuis 2000). However, when consensus does not exist regarding a salient category, social pressures to report to one channel versus another may be diminished. A third scope condition has to do with the severity of the crime, as whistleblowing research shows that severity tends to increase the likelihood of blowing the whistle (Andon et al. 2018; Latan, Jabbour, and de Sousa Jabbour 2021; Tarling and Morris 2010). It is possible that extreme crimes lead to high levels of reporting in all four quadrants and that this reporting is primarily made to external authorities. Indeed, there is some evidence that more severe crimes tend to be reported externally (Callahan and Dworkin 1994; Miceli

et al. 1991). However, there is reason to believe that extreme crimes are not outside the scope of the theory. The sexual abuse of children is widely considered to be an egregious crime, yet the parents of victims almost exclusively reported internally to church officials when the crimes were perpetrated by Catholic priests (Boston Globe 2003). And the crimes examined in this study were certainly serious, yet they still manifest the hypothesized effects. This suggests that even severe forms of wrongdoing may not be exempt from this pattern, but additional study is needed.

Our theory also assumes that individuals are aware of the prevailing views of others within their social environment. In general, we believe this assumption to be reasonable. Not only are people typically aware of the attitudes and opinions of those around them (Dannals, Reit, and Miller 2020; Hall, Mast, and West 2016), but individuals often consult with others before deciding whether or not to report a crime (Greenberg and Ruback 1992). In the latter case, even those ignorant of prevailing attitudes are likely to be exposed to that information through conversations. However, there are certain conditions—such as pluralistic ignorance (Prentice and Miller 1996)—where individuals may be unaware or even misinformed about the views of others. In this case, we expect disapproval or retaliation to manifest subsequent to the reporting decision, which will help individuals to better appraise group attitudes and respond accordingly in the future.

It is also important to acknowledge that in any given context, internal and external authorities may differ in a variety of ways. Already discussed is their access to different punitive methods, but they may also differ in terms of evidentiary standards and their ability to maintain reporter anonymity, among other things. Such differences may affect reporting; for example, an internal authority may have lower evidentiary standards, meaning that witnesses with little evidence are more likely to find support if they report internally. Although we expect such differences to have some impact on whether or not an individual reports internally or externally, we nevertheless expect our theory to hold. Social pressures regarding category affiliation are likely to affect reporting in the theorized manner, net of other factors.

This work provides several opportunities for future research. For one, although the survey is unique in its scope and its ability to test our hypotheses, accounting for the network of attitudes and relationships within a given community (rather than using a single measure of group attitudes) would be beneficial. Furthermore, although we expect the social dynamics observed in Afghan villages to similarly affect reporting behavior in other contexts, future research is needed to establish the theory in other settings. Finally, this work raises a number of questions. For example, what are all the ways in which misalignment comes about? What leads some individuals to report despite being misaligned with their peers? And what leads some individuals to report externally despite being in internal alignment with their community? Answering these and other questions will help to develop the theory further and better clarify the ways in which reporting crimes is not just a personal decision, but a social decision as well.

Notes

1 Although these different authorities may choose to communicate with one another, their systems of justice are functionally self-contained.

2 Of course, individuals are not limited to reporting to only one set of social control agents when multiple outlets are available. Some individuals go to law enforcement in addition to notifying organizational authorities. Most people who go to multiple authorities, however, only do so if the first authority's response is unsatisfactory (Rehg et al. 2004).

3 These numbers should be interpreted with caution, however, as their sample contains only 64 people who reported sexual victimization.

4 Note that we use the word *group* to refer to the organization, community, or family within which internal formal social control takes place. Categories represent classifications into types of people, which may or may not cut across group boundaries. "A social group can be considered as a 'dynamic whole' or social system, characterized by the perceived interdependence among its members, whereas a social category can be defined as a collection of individuals who share at least one attribute in common" (Rabbie and Horwitz 1988:117).

5 Not least of which is the withdrawal and subsequent recapture of Afghanistan by the Taliban while this article was being written. Note that the data used in this study come from 2017 and 2018, when the Taliban controlled four percent of the country and had a presence in—but did not control—an additional 66 percent (Sharifi and Adamou 2018).

6 This may seem surprising considering how quickly the Taliban retook control of Afghanistan in 2021. However, the collapse of Afghanistan is widely attributed to demoralization (due to the U.S. withdrawal) rather than capability. Afghan soldiers vastly outnumbered the Taliban and had far more technologically advanced weaponry (Boot 2021).

7 One exception is Kabul district (the capital), which was subdivided into additional survey units due to the size of the city, which accounts for roughly 13 percent of the country's population. For consistency, we use the administrative boundary designation for Kabul rather than these subdivisions.

8 Interestingly, responses suggest that state capacity was not a major concern in individuals' decisions to report illegal Taliban behavior. Of those who witnessed lawbreaking and did not report the behavior, only 11 percent of respondents explained that they "Don't think security forces will do anything" as the reason for their behavior.

9 Monthly income was reported on a seven-item scale, with the following categories: 2,000 Afghanis or less; 2,001 to 5,000; 5,001 to 10,000; 10,001 to 20,000; 20,001 to 30,000; 30,001 to 40,000; greater than 40,000 Afghanis.

10 Although zakat typically refers to payments made under Islamic law for charitable and religious purposes, zakat in this context refers specifically to illegal and extortionary taxation imposed by the Taliban.

11 Randomization leads any subsequent measurement error to be classical, which is downwardly biased.

12 If this cooperation is widely known, however, it might make individuals reticent to report internally, as the information reported would no longer be confined within the group.

References

Akerstrom, Malin. 1988. "The Social Construction of Snitches." *Deviant Behavior* 9(2):155–67. https://doi.org/10.1080/01639625.1988.9967776.

Alekseyeva, Kristina, Stephanie Birndorf, Joseph DeMott, Shannon Galvin, Mehdi J. Hakimi, Juan Pablo Perez-Sangimino, Tara Rangchi, and Michael Talent. 2017. *An Introduction to the Law of Afghanistan.* Stanford, CA: Stanford Law School, Afghanistan Legal Education Project, 4th edition.

Alexander, Jeffrey C. 2018. "The Societalization of Social Problems: Church Pedophilia, Phone Hacking, and the Financial Crisis." *American Sociological Review* 83(6):1049–78. https://doi.org/10.1177/0003122418803376.

Andon, Paul, Clinton Free, Radzi Jidin, Gary S. Monroe, and Michael J. Turner. 2018. "The Impact of Financial Incentives and Perceptions of Seriousness on Whistleblowing Intention." *Journal of Business Ethics* 51(1):165–78. https://doi.org/10.1007/s10551-016-3215-6.

Aven, Brandy L. 2015. "The Paradox of Corrupt Networks: An Analysis of Organizational Crime at Enron." *Organization Science* 26(4):980–96. https://doi.org/10.1287/orsc.2015.0983.

Baldassarri, Delia, and Guy Grossman. 2011. "Centralized Sanctioning and Legitimate Authority Promote Cooperation in Humans." *Proceedings of the National Academy of Sciences* 108(27):11023–27. https://doi.org/10.1073/pnas.1105456108.

Barfield, Thomas. 2013. "Continuities and Changes in Local Politics in Northern Afghanistan." In *Local Politics in Afghanistan: A Century of Intervention in the Social Order*, edited by C. Schetter. London: Hurst and Company.

Barnett, Tim, Daniel S. Cochran, and G. Stephen Taylor. 1993. "The Internal Disclosure Policies of Private-Sector Employers: An Initial Look at Their Relationship to Employee Whistleblowing." *Journal of Business Ethics* 12:127–36. https://doi.org/10.1007/BF00871932.

Beal, James F. 2016. *Mission Accomplished? Rebuilding the Iraqi and Afghan Armies.* Master's thesis, Naval Postgraduate School, Monterey, CA.

Bergemann, Patrick. 2019. *Judge Thy Neighbor: Denunciations in the Spanish Inquisition, Romanov Russia, and Nazi Germany.* Columbia University Press.

Bergemann, Patrick, and Brandy Aven. 2022. "Whistleblowing and Group Affiliation: The Role of Group Cohesion and the Locus of the Wrongdoer in Reporting Decisions." *Organization Science,* first published August 12, 2022. https://doi.org/10.1287/orsc.2022.1617.

Berman, Eli, Michael Callen, Joseph H. Felter, and Jacob N. Shapiro. 2011. "Do Working Men Rebel? Insurgency and Unemployment in Afghanistan, Iraq, and the Philippines." *Journal of Conflict Resolution* 55(4):496–528. https://doi.org/10.1177/0022002710393920.

Black, Donald J. 1976. *The Behavior of Law*. Bingley, UK: Emerald Group.

Blair, Graeme, Kosuke Imai, and Jason Lyall. 2014. "Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan." *American Journal of Political Science* 58(4):1043–63. https://doi.org/10.1111/ajps.12086.

Boot, Max. 2021. "How the Afghan Army Collapsed Under the Taliban's Pressure." Council on Foreign Relations. https://www.cfr.org/in-brief/how-afghan-army-collapsed-under-talibans-pressure.

Boston Globe. 2003. *Betrayal: The Crisis in the Catholic Church*. Boston: Back Bay.

Brown, A. J. (ed.). 2008. *Whistleblowing in the Australian Public Sector: Enhancing the Theory and Practice of Internal Witness Management in Public Sector Organisations*. Canberra, Australia: ANU Press.

Callahan, Elletta Sangrey, and Terry Morehead Dworkin. 1994. "Who Blows the Whistle to the Media, and Why: Organizational Characteristics of Media Whistle-blowers." *American Business Law Journal* 32(2):151–84. https://doi.org/10.1111/j.1744-1714.1994.tb00933.x.

Chen, Ching-Pu, and Chih-Tsung Lai. 2014. "To Blow or Not to Blow the Whistle: The Effects of Potential Harm, Social Pressure and Organisational Commitment on Whistleblowing Intention and Behaviour." *Business Ethics: A European Review* 23(3):327–42. https://doi.org/10.1111/beer.12053.

Cheng, Jin, Haiqing Bai, and Xijuan Yang. 2019. "Ethical Leadership and Internal Whistleblowing: A Mediated Moderation Model." *Journal of Business Ethics* 155(1):115–30. https://doi.org/10.1007/s10551-017-3517-3.

Child, Travers, Austin L. Wright, and Yun Xiao. 2021. "Aid Fragmentation and Corruption." Becker Friedman Institute for Economics Working Paper No. 2021-69.

Child, Travers Barclay. 2019. "Conflict and Counterinsurgency Aid: Drawing Sectoral Distinctions." *Journal of Development Economics* 141:102245. https://doi.org/10.1016/j.jdeveco.2018.06.003.

Coburn, Noah. 2011. *Bazaar Politics: Power and Pottery in an Afghan Market Town*. Stanford University Press.

Cohen, Stanley. 1985. *Visions of Social Control: Crime, Punishment, and Classification*. Cambridge: Polity Press.

Condra, Luke N., James D. Long, Andrew C. Shaver, and Austin L. Wright. 2018. "The Logic of Insurgent Electoral Violence." *American Economic Review* 108(11):3199–231. https://doi.org/10.1257/aer.20170416.

Condra, Luke N., and Austin L. Wright. 2019. "Civilians, Control, and Collaboration during Civil Conflict." *International Studies Quarterly* 63(4):897–907. https://doi.org/10.1093/isq/sqz042.

Cortina, Lilia M., and Vicki J. Magley. 2003. "Raising Voice, Risking Retaliation: Events Following Interpersonal Mistreatment in the Workplace." *Journal of Occupational Health Psychology* 8(4):247–65. https://doi.org/10.1037/1076-8998.8.4.247.

Crisp, Richard J., and Miles Hewstone. 2007. "Multiple Social Categorization." *Advances in Experimental Social Psychology* 39:163–254. https://doi.org/10.1016/S0065-2601(06)39004-1.

Culiberg, Barbara, and Katarina Katja Mihelič. 2017. "The Evolution of Whistleblowing Studies: A Critical Review and Research Agenda." *Journal of Business Ethics* 146(4):787–803. https://doi.org/10.1007/s10551-016-3237-0.

Dannals, Jennifer E., Emily S. Reit, and Dale T. Miller. 2020. "From Whom Do We Learn Group Norms? Low-Ranking Group Members Are Perceived as the Best Sources." *Organizational Behavior and Human Decision Processes* 161:213–27. https://doi.org/10.1016/j.obhdp.2020.08.002.

Donkin, Marika, Rodney Smith, and Arthur J. Brown. 2008. "How Do Officials Report? Internal and External Whistleblowing." In *Whistleblowing in the Australian Public Sector: Enhancing the Theory and Practice of Internal Witness Management in Public Sector Organisations*, edited by A. J. Brown. Canberra, Australia: ANU Press.

Dozier, Janelle Brinker, and Marcia P. Miceli. 1985. "Potential Predictors of Whistle-Blowing: A Prosocial Behavior Perspective." *Academy of Management Review* 10(4):823–36. https://doi.org/10.5465/amr.1985.4279105.

Drori, Israel, Amy Wrzesniewski, and Shmuel Ellis. 2013. "One out of Many? Boundary Negotiation and Identity Formation in Postmerger Integration." *Organization Science* 24(6):1717–41. https://doi.org/10.1287/orsc.1120.0814.

Dungan, James A., Liane Young, and Adam Waytz. 2019. "The Power of Moral Concerns in Predicting Whistleblowing Decisions." *Journal of Experimental Social Psychology* 85:103848. https://doi.org/10.1016/j.jesp.2019.103848.

Dworkin, Terry Morehead, and Melissa S. Baucus. 1998. "Internal vs. External Whistleblowers: A Comparison of Whistleblowering Processes." *Journal of Business Ethics* 17(12):1281–98. https://doi.org/10.1023/A:1005916210589.

Eisenhardt, Kathleen M. 1989. "Building Theories from Case Study Research." *Academy of Management Review* 14(4):532–50. https://doi.org/10.5465/amr.1989.4308385.

Ellickson, Robert C. 1987. "A Critique of Economic and Sociological Theories of Social Control." *Journal of Legal Studies* 16(1):67–99. https://doi.org/10.1086/467824.

Erikson, Kai. 1966. *Wayward Puritans: A Study in the Sociology of Deviance.* New York: John Wiley & Sons.

Farrell, Theo, and Antonio Giustozzi. 2013. "The Taliban at War: Inside the Helmand Insurgency, 2004–2012." *International Affairs* 89(4):845–71. https://doi.org/10.1111/1468-2346.12048.

Fetzer, Thiemo, Pedro C. L. Souza, Oliver Vanden Eynde, and Austin L. Wright. 2021. "Security Transitions." *American Economic Review* 111(7):2275–308. https://doi.org/10.1257/aer.20200412.

Fiorin, Stefano. 2021. "Reporting Peers' Wrongdoing: Experimental Evidence on the Effect of Financial Incentives on Morally Controversial Behavior." Working paper.

Fisher, Bonnie S., Leah E. Daigle, Francis T. Cullen, and Michael G. Turner. 2003. "Reporting Sexual Victimization to the Police and Others: Results from a National-Level Study of College Women." *Criminal Justice and Behavior* 30(1):6–38. https://doi.org/10.1177/0093854802239161.

Gao, Jingyu, Robert Greenberg, and Bernard Wong-On-Wing. 2015. "Whistleblowing Intentions of Lower-Level Employees: The Effect of Reporting Channel, Bystanders, and Wrongdoer Power Status." *Journal of Business Ethics* 126(1):85–99. https://doi.org/10.1007/s10551-013-2008-4.

Giustozzi, Antonio. 2015. *The Army of Afghanistan: A Political History of a Fragile Institution*. London: Hurst.

Gökçe, Asiye Toker. 2013. "Teachers' Value Orientations as Determinants of Preference for External and Anonymous Whistleblowing." *International Journal of Humanities and Social Science* 3(4):163–73.

Greenberg, Martin S., and R. Barry Ruback. 1992. *After the Crime: Victim Decision Making*. New York: Springer.

Habib, Shabnam. 2013. "Local Government in Afghanistan: How It Works and Main Challenges." In *Proceedings of the 21th NISPAcee Annual Conference*.

Hall, Judith A., Marianne Schmid Mast, and Tessa V. West. 2016. *The Social Psychology of Perceiving Others Accurately*. Cambridge University Press.

Heckman, James. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1):153–61. https://doi.org/10.2307/1912352.

Henderson-King, Eaaron I., and Richard E. Nisbett. 1996. "Anti-Black Prejudice as a Function of Exposure to the Negative Behavior of a Single Black Person." *Journal of Personality and Social Psychology* 71(4):654–64. https://doi.org/10.1037/0022-3514.71.4.654.

Hernes, Tor. 2004. "Studying Composite Boundaries: A Framework of Analysis." *Human Relations* 57(1):9–29. https://doi.org/10.1177/0018726704042712.

Hogg, Michael A. 2004. "Social Categorization, Depersonalization, and Group Behavior." In *Blackwell Handbook of Social Psychology: Group Processes*, edited by M. A. Hogg and R. S. Tindale, pp. 56–85. Malden, MA: Blackwell Publishers.

Hogg, Michael A., and John C. Turner. 1987. "Intergroup Behaviour, Self-Stereotyping and the Salience of Social Categories." *British Journal of Social Psychology* 26(4):325–40. https://doi.org/10.1111/j.2044-8309.1987.tb00795.x.

Hollinger, Richard C., and John P. Clark. 1982. "Formal and Informal Social Controls of Employee Deviance." *Sociological Quarterly* 23(3):333–43. https://doi.org/10.1111/j.1533-8525.1982.tb01016.x.

Hunt, Scott A., and Robert D. Benford. 2004. "Collective Identity, Solidarity, and Commitment." In *The Blackwell Companion to Social Movements*, edited by D. A. Snow, S. A. Soule, and H. Kriesi, pp. 433–57. Malden, MA: Blackwell Publishing.

Jeon, So Hee. 2017. "Where to Report Wrongdoings? Exploring the Determinants of Internal versus External Whistleblowing." *International Review of Public Administration* 22(2):153–71. https://doi.org/10.1080/12294659.2017.1315235.

Jubb, Peter B. 1999. "Whistleblowing: A Restrictive Definition and Interpretation." *Journal of Business Ethics* 21(1):77–94. https://doi.org/10.1023/A:1005922701763.

Kalyvas, Stathis. 2006. *The Logic of Violence in Civil War*. Cambridge University Press.

Kaptein, Muel. 2011. "From Inaction to External Whistleblowing: The Influence of the Ethical Culture of Organizations on Employee Responses to Observed Wrongdoing." *Journal of Business Ethics* 98(3):513–30. https://doi.org/10.1007/s10551-010-0591-1.

Karell, Daniel, and Michael Freedman. 2019. "Rhetorics of Radicalism." *American Sociological Review* 84(4):726–53. https://doi.org/10.1177/0003122419859519.

Katzman, Kenneth. 2011. "Afghanistan: Politics, Elections, and Government Performance." Congressional Research Service (CRS) Report RS21922.

Keblusek, Lauren, Howard Giles, and Anne Maass. 2017. "Communication and Group Life: How Language and Symbols Shape Intergroup Relations." *Group Processes & Intergroup Relations* 20(5):632–43. https://doi.org/10.1177/1368430217708864.

Keenan, Marie. 2013. *Child Sexual Abuse and the Catholic Church: Gender, Power, and Organizational Culture*. Oxford University Press.

Langley, Ann, Kajsa Lindberg, Bjørn Erik Mørk, Davide Nicolini, Elena Raviola, and Lars Walter. 2019. "Boundary Work among Groups, Occupations, and Organizations: From Cartography to Process." *Academy of Management Annals* 13(2):704–36. https://doi.org/10.5465/annals.2017.0089.

Latan, Hengky, Charbel Jose Chiappetta Jabbour, and Ana Beatriz Lopes de Sousa Jabbour. 2021. "To Blow or Not to Blow the Whistle: The Role of Rationalization in the Perceived Seriousness of Threats and Wrongdoing." *Journal of Business Ethics* 169(3):517–35. https://doi.org/10.1007/s10551-019-04287-5.

Lee, Gladys, and N. L. Fargher. 2014. "When Does Financial Misconduct Result in External Whistle-Blowing: A Comparison between External Whistle-Blowing and Internal Reporting." Unpublished manuscript.

Light, Michael T., and John Iceland. 2016. "The Social Context of Racial Boundary Negotiations: Segregation, Hate Crime, and Hispanic Racial Identification in Metropolitan America." *Sociological Science* 3:61–84. https://doi.org/10.15195/v3.a4.

Meier, Robert F. 2018. "Deviance, Social Control, and Criminalization." In *The Handbook of Social Control*, edited by M. Deflem. Hoboken, NJ: John Wiley & Sons.

Mesmer-Magnus, Jessica R., and Chockalingam Viswesvaran. 2005. "Whistle-blowing in Organizations: An Examination of Correlates of Whistleblowing Intentions, Actions, and Retaliation." *Journal of Business Ethics* 62(3):277–97. https://doi.org/10.1007/s10551-005-0849-1.

Miceli, Marcia P., Janet Pollex Near, and Terry M. Dworkin. 2008. *Whistle-Blowing in Organizations*. New York: Routledge.

Miceli, Marcia. P., Michael Rehg, Janet Near, and Katherine Ryan. 1999. "Can Laws Protect Whistle-Blowers? Results of a Naturally Occurring Field Experiment." *Work and Occupations* 26(1):129–51. https://doi.org/10.1177/0730888499026001007.

Miceli, Marcia P., and Janet P. Near. 2002. "What Makes Whistle-Blowers Effective? Three Field Studies." *Human Relations* 55(4):455–79. https://doi.org/10.1177/0018726702055004463.

Miceli, Marcia P., Janet P. Near, and Charles R. Schwenk. 1991. "Who Blows the Whistle and Why?" *Industrial and Labor Relations Review* 45(1):113–30. https://doi.org/10.1177/001979399104500108.

Miethe, Terance D. 1999. *Whistleblowing at Work: Tough Choices in Exposing Fraud, Waste, and Abuse on the Job*. Boulder, CO: Westview Press.

Murtazashvili, Jennifer Brick. 2016. *Informal Order and the State in Afghanistan*. Cambridge University Press.

Nayir, Dilek Zamantili, and Christian Herzig. 2012. "Value Orientations as Determinants of Preference for External and Anonymous Whistleblowing." *Journal of Business Ethics* 107(2):197–213. https://doi.org/10.1007/s10551-011-1033-4.

Nayir, Dilek Zamantili, Michael T. Rehg, and Yurdanur Asa. 2018. "Influence of Ethical Position on Whistleblowing Behaviour: Do Preferred Channels in Private and Public Sectors Differ?" *Journal of Business Ethics* 149(1):147–67. https://doi.org/10.1007/s10551-016-3035-8.

Orchowski, Lindsay M., Douglas H. Meyer, and Christine A. Gidycz. 2009. "College Women's Likelihood to Report Unwanted Sexual Experiences to Campus Agencies: Trends and Correlates." *Journal of Aggression, Maltreatment and Trauma* 18(8):839–58. https://doi.org/10.1080/10926770903291779.

Pain, Adam. 2016. "Using Village Context Analysis in Afghanistan: Methods and Wider Implications." Secure Livelihoods Research Consortium (SLRC) Working Paper 47.

Palmer, Donald. 2012. *Normal Organizational Wrongdoing: A Critical Analysis of Theories of Misconduct in and by Organizations*. Oxford University Press.

Palmer, Donald, and Valerie Feldman. 2018. *Comprehending the Incomprehensible: Organization Theory and Child Sexual Abuse in Organizations*. Boulder, CO: Westview Press.

Park, Heungsik, John Blenkinsopp, M. Kemal Oktem, and Ugur Omurgonulsen. 2008. "Cultural Orientation and Attitudes toward Different Forms of Whistle-blowing: A Comparison of South Korea, Turkey, and the UK." *Journal of Business Ethics* 82(4):929–39. https://doi.org/10.1007/s10551-007-9603-1.

Paul, Lisa A., Heidi M. Zinzow, Jenna L. McCauley, Dean G. Kilpatrick, and Heidi S. Resnick. 2014. "Does Encouragement by Others Increase Rape Reporting? Findings from a National Sample of Women." *Psychology of Women Quarterly* 38(2):222–32. https://doi.org/10.1177/0361684313501999.

Pettigrew, Andrew M. 1995. "Longitudinal Field Research on Change: Theory and Practice." In *Longitudinal Field Research Methods: Studying Processes of Organizational Change*, edited by G. P. Huber and A. H. Van den Ven. Thousand Oaks, CA: SAGE Publications.

Piazza, Alessandro, and Julien Jourdan. 2018. "When the Dust Settles: The Consequences of Scandals for Organizational Competition." *Academy of Management Journal* 61(1):165–90. https://doi.org/10.5465/amj.2015.1325.

Podles, Leon J. 2008. *Sacrilege: Sexual Abuse in the Catholic Church*. Baltimore, MD: Crossland Press.

Prentice, Deborah A., and Dale T. Miller. 1996. "Pluralistic Ignorance and the Perpetuation of Social Norms by Unwitting Actors." *Advances in Experimental Social Psychology* 28:161–209. https://doi.org/10.1016/S0065-2601(08)60238-5.

Rabbie, Jacob M., and Murray Horwitz. 1988. "Categories versus Groups as Explanatory Concepts in Intergroup Relations." *European Journal of Social Psychology* 18(2):117–23. https://doi.org/10.1002/ejsp.2420180204.

Rehg, Michael. 1998. *An Examination of the Retaliation Process against Whistleblowers: A Study of Federal Government Employees*. Ph.D. thesis, Indiana University.

Rehg, Michael T., Marcia P. Miceli, Janet P. Near, and James R. Van Scotter. 2004. "Predicting Retaliation against Whistle-Blowers: Outcomes of Power Relationships within Organizations." In *Academy of Management Proceedings*. Briarcliff Manor, NY: Academy of Management. https://doi.org/10.5465/ambpp.2004.13857462.

Rhodes, Marjorie, and Andrew Baron. 2019. "The Development of Social Categorization." *Annual Review of Developmental Psychology* 1:359–86. https://doi.org/10.1146/annurev-devpsych-121318-084824.

Rothschild, Joyce, and Terance D. Miethe. 1999. "Whistle-Blower Disclosures and Management Retaliation: The Battle to Control Information about Organization Corruption." *Work and Occupations* 26(1):107–28. https://doi.org/10.1177/0730888499026001006.

Ruback, R. Barry, Martin S. Greenberg, and David R. Westcott. 1984. "Social Influence and Crime-Victim Decision Making." *Journal of Social Issues* 40(1):51–76. https://doi.org/10.1111/j.1540-4560.1984.tb01082.x.

Sharifi, Shoaib, and Louise Adamou. 2018. "Taliban Threaten 70% of Afghanistan, BBC Finds." *BBC News,* January 31. https://www.bbc.com/news/world-asia-42863116.

Shaver, Andrew, and Jacob N. Shapiro. 2021. "The Effect of Civilian Casualties on Wartime Informing: Evidence from the Iraq War." *Journal of Conflict Resolution* 65(7–8):1337–77. https://doi.org/10.1177/0022002721991627.

Sims, Randi L., and John P. Keenan. 1998. "Predictors of External Whistleblowing: Organizational and Intrapersonal Variables." *Journal of Business Ethics* 17(4):411–21. https://doi.org/10.1023/A:1005763807868.

Sonin, Konstantin, and Austin L. Wright. 2022. "Information Operations Increase Civilian Security Cooperation." *Economic Journal* 132(643):1179–99. https://doi.org/10.1093/ej/ueab101.

Strand, Arne. 2013. "Perspectives of Local Violence: Revenge, Mediation and Conflict Resolution." In *Local Politics in Afghanistan: A Century of Intervention in the Social Order*, edited by C. Schetter. London: Hurst and Company.

Tarling, Roger, and Katie Morris. 2010. "Reporting Crime to the Police." *British Journal of Criminology* 50(3):474–90. https://doi.org/10.1093/bjc/azq011.

Vadera, Abhijeet K., Ruth V. Aguilera, and Brianna B. Caza. 2009. "Making Sense of Whistle-Blowing's Antecedents: Learning from Research on Identity and Ethics Programs." *Business Ethics Quarterly* 19(4):553–86. https://doi.org/10.5840/beq200919432.

Van Knippenberg, Ad, and Ap Dijksterhuis. 2000. "Social Categorization and Stereotyping: A Functional Perspective." *European Review of Social Psychology* 11(1):105–144. https://doi.org/10.1080/14792772043000013.

Wardak, Ali. 2011. "State and Non-state Justice Systems in Afghanistan: The Need for Synergy." *University of Pennsylvania Journal of International Law* 32(5):1305–24.

Wardak, Ali, and John Braithwaite. 2012. "Crime and War in Afghanistan. Part II: A Jeffersonian Alternative?" *British Journal of Criminology* 53(2):197–214. https://doi.org/10.1093/bjc/azs066.

Wardak, Ali, Daud Saba, and Halima Kazem. 2007. "Bridging Modernity and Tradition: Rule of Law and the Search for Justice." Afghanistan Human Development Report 2007, Center for Policy and Human Development at Kabul University.

Wimmer, Andreas. 2007. "How (Not) to Think about Ethnicity in Immigrant Societies." ESRC Centre on Migration, Policy and Society Working Paper No. 44, Oxford University.

Wootson, Cleve. 2018. "A Black Yale Student Fell Asleep in Her Dorm's Common Room. A White Student Called Police." *Washington Post,* May 10. https://www.washingtonpost.com/news/grade-point/wp/2018/05/10/a-black-yale-student-fell-asleep-in-her-dorms-common-room-a-white-student-called-police.

Zipay, Kate P., Marie S. Mitchell, Michael D. Baer, Hudson Sessions, and Robert J. Bies. 2021. "Lenient Reactions to Misconduct: Examining the Self-Conscious Process of Being Lenient to Others at Work." *Academy of Management Journal* 64(2):351–77. https://doi.org/10.5465/amj.2018.0123.

Zipparo, Lisa. 1999. "Encouraging Public Sector Employees to Report Workplace Corruption." *Australian Journal of Public Administration* 58(2):83–93. https://doi.org/10.1111/1467-8500.00091.

Patrick Bergemann: Paul Merage School of Business, University of California, Irvine. E-mail: pbergema@uci.edu.

Austin L. Wright: Harris School of Public Policy, The University of Chicago. E-mail: austinlw@uchicago.edu.