
Stopping the Violence but Blocking the Peace: Dilemmas of Foreign-Imposed Nation Building After Ethnic War

Kevin Russell and Nicholas Sambanis 

Abstract Can third parties build nations after ethno-sectarian war? We provide a positive theory of peace building that highlights trade-offs that are inherent in any foreign intervention, narrowing the conditions for success even when interventions are well resourced and even-handed. A “sectarian” dilemma arises because peace must rely on local leaders, but leaders who earned their reputations through ethno-sectarian conflict have no incentive to stop playing the ethnic card and will not provide public goods. Intervention can shift those incentives if it stops ethnic violence and rebuilds state institutions. But an “institutional” dilemma arises if intervention crowds out local leaders, limiting state legitimacy and constraining the pace with which state building unfolds. The window for a lengthier, slower pace of foreign-led state building will close due to its own success as the population switches from ethnic to national identification, creating a “sovereignty” dilemma that pushes third parties out. If intervention ends before institutions can deepen leader incentives for a unifying nationalism, violence will likely recur. We provide an “intervention diagnostic” that reflects these three dilemmas, which are a function of the type of intervention, local political development, and the identity of the intervener. In deciding whether to intervene, the limits of building self-enforcing peace should be weighed against the likelihood and costs of ongoing violence.

In March 2011, pro-democracy protests in Syria’s southern city of Daraa quickly escalated into civil war along ethno-sectarian lines. After eight years of war, twelve million people had been displaced and 400,000 had been killed. American President Barack Obama was criticized for not intervening to stop the fighting. But the US-led intervention in Libya in 2011 had been motivated by just such a desire to protect civilians and left the country fragmented along tribal lines. Critics argued that the Libyan intervention was too limited, forcibly removing the regime while leaving a power vacuum that inevitably led to civil war. Yet if the American footprint in Libya was too small it reflected a wariness of heavy-handed intervention after the prolonged American occupation in Iraq failed in its mandate to unify the Iraqi nation, despite US military spending of USD 800 billion and a peak force of 171,000 troops.

The international community’s intolerance of mass atrocities and civilian suffering in war has increased, leading to calls for more interventions to end ethnic wars before the parties have fought to a stalemate. Can such interventions succeed? We provide a new positive theory of the promise and perils of third-party intervention in ethnic war and set new terms for a debate that has so far been framed by pulls in two directions.

Interventionists argue that humanitarian catastrophe mandates intervention, yet a limited military response is often not sufficient to prevent the recurrence of violent conflict. The “responsibility to protect” also implies a responsibility to follow through, and some form of transitional administration will usually be necessary to prevent a violent struggle where ethnic identity readily evokes existential conflicts in politics.¹ Indeed, the risk of war recurrence is high after civil war; more than a third of all terminated civil wars recur within five years.² Because light-footprint interventions that are limited to the provision of humanitarian assistance cannot resolve the underlying causes of conflict, more intrusive interventions are required. Peacekeeping can help, but only when the parties are ready to settle. The imperative to intervene early to end civilian suffering pushes interventions to become increasingly “holistic” by promoting state building, power sharing, and nation building as solutions to ethnic strife.

Skeptics point to mounting evidence against the effectiveness of such interventions. Peace building increasingly involves democracy promotion, but democracy is a distant goal for all but a few postwar countries.³ Limited interventions targeting regime change are rarely successful.⁴ Transitional administration or occupation appears necessary to help the weakest states regain their footing, but heavy-handed interventions are costly and will face an “obsolescing welcome.”⁵ Light-footprint interventions might generate less backlash, but they could result in a power vacuum that promotes new conflicts.⁶ These challenges have led many to question the very logic of state-building intervention, because it often produces fragile outcomes subject to reversals.⁷ These arguments are consistent with an American public wary of US intervention in anything beyond a counterterrorism mission as enduring peace appears elusive across Iraq, Afghanistan, Syria, and Libya.

We contribute to this debate by developing a new theory of third-party-led state building that explains when and how intervention in ethno-sectarian wars can be effective in building self-enforcing peace.⁸ A key focus of our theory is on building peace by unifying a nation and shifting leaders’ political incentives to work within national institutions rather than cater to ethnic constituencies. These mechanisms help illuminate the conditions for success, but also reveal limits on third-party effectiveness.

There are three main obstacles to building peace after ethnic war—three interlocking dilemmas that are inherent in the logic of the new interventionism. A *sectarian*

1. On the evolution of the “responsibility to protect” doctrine, see the UN Office on Genocide Prevention and the Responsibility to Protect, <<https://www.un.org/en/genocideprevention/about-responsibility-to-protect.shtml>>.

2. Doyle and Sambanis 2006.

3. Brownlee 2007.

4. Downes and Monten 2013.

5. Edelstein 2009.

6. Kuperman 2013; Paris 2004.

7. Sisk 2013, 166.

8. We use the terms *ethnic*, *sectarian*, and *ethno-sectarian* interchangeably. Our theory applies broadly to any group identity that has the features that we ascribe to groups in our model.

dilemma arises when a peace agreement in the aftermath of war must rely on a bargain among local leaders, warlords, or power brokers. Leaders who have cultivated ethno-sectarian constituencies will have no incentive to provide public goods and build national institutions. Power sharing among such leaders will reify the very ethnic divisions of the conflict. The dilemma that arises is whether to intervene with a light touch, enabling power sharing and a quick third-party exit, while risking the reification of sectarianism; or to be heavy-handed, intervening forcefully and at greater cost, but to thereby reduce ethnic violence, transform the conflict, and take more control of the process of rebuilding national institutions so as to create space for the population to shift from ethnic to national identification. If the population begins to identify nationally, this will reduce the appeal of ethnic agendas and will enable local leaders to cater to constituencies outside their ethnic or sectarian group.

The sectarian dilemma appears as a test of will for the intervener, which is often how the debate about intervention is framed: is the intervener (or the international community) willing to commit the lives and resources necessary to transform the very nature of the conflict through state building and nation building? Or are they looking for a fast way out? However, framing the debate in this way is incomplete because under some conditions heavy-handed intervention can backfire.

Third-party efforts to reduce ethnic violence and rebuild the state will not create self-enforcing peace if the intervention crowds out local leaders, weakens their credibility vis-à-vis political audiences, and removes their incentives to support state building. This *institutional dilemma* lengthens the time needed to build national institutions. At the same time, the window of opportunity for success begins to close as the population shifts from ethnic to national identification, if nationalism is perceived as antithetical to foreign rule. This *sovereignty dilemma* shortens the time horizon for intervention if nationalism induces leaders seeking unity to push the intervener out. The sovereignty dilemma is particularly challenging for the nation-building path to peace because it is induced by the initial success of that very strategy. These interacting dilemmas impose limits on the pace and duration of intervention, and external assistance for state building could end before institutions are developed to the point where they can constrain ethno-sectarian elites. This sets the stage for ethnic violence to recur.

We explore these insights in a theory that is based on local politics. A key insight is that the short-term provision of technical capacity and development assistance is not enough to ensure that conflicts will be resolved. Institutions must gain credibility through use; and interventions will succeed when they incentivize leaders to operate through state institutions consistently over time, proving their commitment to a constitutional plan that citizens and interest groups will come to rely on and thus defend, thereby creating self-enforcing peace. But there is an inherent contradiction in external forces trying to strengthen the link between leaders and the public and trying to confer legitimacy on the unified state, giving rise to the institutional and sovereignty dilemmas, which constrain the possibilities for peace-building success. Arguments for third parties to commit the necessary resources to overcome the sectarian dilemma may set the bar for intervention too low if they do not account for these complications.

Our theory stands in sharp contrast to other major contributions to the literature on peace building. Many recent studies are limited to criticizing different strategies of intervention,⁹ and they do not develop a positive theory of the conditions under which intervention can succeed. We go beyond identifying bias or resource constraints as the source of intervention failure and consider the limits of even well-resourced interventions that are designed to build an inclusive state. We draw on a psychological framework for how group identities respond to conflict, and we show that inter-ethnic cooperation *is* possible after ethnic war under some conditions. Contrary to views that partition is the only way to resolve ethnic conflicts after they become very violent,¹⁰ our theory shows that self-enforcing peace among non-co-ethnics can be achieved via nation building. Contrary to views that the settlement of civil wars is only possible with third-party enforcement to resolve commitment problems,¹¹ we show how transformation of the population's social identities can incentivize leaders to invest in national institutions even after external guarantees are withdrawn. We also go beyond previous approaches to peacekeeping and peace building¹² by emphasizing the role of local leaders in solidifying the population's shift from ethno-sectarian to national identities. Previous scholarship has demonstrated the positive effects of multilateral (especially United Nations) operations in protecting civilians, holding elections, training police, and helping rebuild the state, but it does not explain how identities shape the process of creating elite incentives for state building and nation building for the long run.

This argument is developed formally with the help of a model that incorporates these causal relationships and identifies parameters that make the three dilemmas more or less severe. We discuss the implications of the model for the prospects of success in different types of peace-building ecologies. These ecologies are shaped by the intensity of ethnic versus national identification; the quality of state institutions; leaders' willingness to provide public goods across ethnic lines; and the intensity of the institutional and sovereignty dilemmas. Different types of intervention are optimal in different peace-building ecologies. We also present our arguments informally, with the help of diagrams, in the next section. The theory is broad and could be applicable to most conflicts and most types of intervention, because it speaks to ongoing public debates about the right way for the international community to intervene in civil wars.

We use our model as a guide to discuss several illustrative examples of foreign intervention in ethnic war, starting with the failure of peace building in Iraq. Iraq's civil war had a huge impact on the politics of the region and has also been pivotal in rethinking US counterinsurgency strategy. Between 2007 and 2009 the United States surged its forces to overcome the sectarian dilemma and induce a nationalist

9. Lake 2016; Paris 2014.

10. Kaufmann 1996.

11. Walter 1997.

12. Doyle and Sambanis 2006.

shift in social identities in Iraq. But American occupation could only control so much or push to build institutions so fast without crowding out Iraqi leaders, who were also wary of being seen as puppets of the United States. This limited the pace of state building at a time when a shift away from sectarianism and toward nationalism raised the specter of the sovereignty dilemma, ultimately forcing the withdrawal of US troops and allowing the return of sectarianism. The theory shows why Iraq's leaders had strong incentives to respond to the public's nationalist shift, but also to anticipate (once the exit plan came into view) that nationalist politics might not hold after the US withdrawal. In the empirical section, we generalize this discussion by outlining a typology of interventions and their anticipated outcomes, including how the identity of the third party might lessen the severity of the sovereignty dilemma and conditions where incremental state building might be superior to the nation-building path to peace.

Theory

A large literature has debated, in a piecemeal fashion, the pros and cons of peace-building interventions. Scholars have asked whether development assistance can provide security benefits;¹³ whether external enforcement can build peace after civil war¹⁴ and whether such peace can be self-sustaining¹⁵ by promoting the rule of law;¹⁶ whether alien rule can succeed if it is effective at providing public goods;¹⁷ and whether the key to successful peace building is centralized control of any intervention,¹⁸ or whether peace building must be based on co-opting existing state institutions in the target country.¹⁹ These and many other complex questions frame an unwieldy policy debate on the effectiveness of intervention. Most studies focus on a single issue and do not consider how solving one problem might generate another.

Scholarship on postwar transitions has not yet addressed the interplay between institutional development, the incentives of political leaders, and the social preferences of the population. We show that the provision of material resources or security alone is not sufficient to build peace, and that peace develops as a result of interventions that can change leader incentives. We look beyond studies of the effects of development assistance in conflict settings to show that the transformation of third-party resources into public goods depends in part on the population's social identification, which determines whether and how leaders will use those resources for state building. In our theory, self-enforcing peace becomes possible through state building;

13. Beath, Christia, and Enikolopov 2012.

14. Edelstein 2009.

15. Doyle and Sambanis 2006; Matanock and Lichtenheld 2015.

16. Blair 2015.

17. Hechter 2013; Lawrence 2013.

18. Fukuyama 2004.

19. Dobbins et al. 2007.

and state building becomes possible via nation building.²⁰ But can nation building be achieved under foreign occupation?

A Nontechnical Presentation of the Argument

Previous scholarship has argued that state-building interventions are likely to fail if the occupier's preferences are different from the population's and an occupier installs a puppet,²¹ or if intervention is under-resourced or narrow.²² But what about cases where the intervener is impartial and not resource- or time-constrained?

We deliberately abstract from considering the third party's objectives or constraints and focus on "best case" scenarios where a well-resourced third party is prepared to invest sufficient resources to end an ethno-sectarian war and reunify the nation. These cases bring into focus the challenges stemming from local politics rather than deficiencies of intervention. We focus on conflicts where ethno-sectarian identities and violence are mutually reinforcing among groups that compete for control of the state and its resources. Peace requires a shift to conditions where a shared national identity and security are mutually reinforcing, but this will be stable only where institutions are strong enough to keep the rewards for violence low and the benefit of public goods high. We are interested in the case where ethnic hostilities are strong enough, and state institutions weak enough, that leaders are unlikely to build national institutions and cater to the needs of non-co-ethnics without outside intervention.

The process by which intervention shapes local politics is represented in [Figure 1](#). At point L1 ethnic groups can appropriate resources by violence, generating conflict that induces the population to identify ethnically. More intense conflict along ethno-sectarian lines makes ethno-sectarian identities more salient. From the perspective of an intervening force, these conditions present the sectarian dilemma.

Sectarian dilemma: After ethno-sectarian war, a peace settlement without major intervention or occupation requires support from ethno-sectarian leaders, but empowering leaders who have secured their reputations looking after ethnic constituencies will not give them incentives to cater to the nation and invest in the state building needed to shift to self-enforcing peace.

A peace agreement brokered based on the balance of power among warring groups incentivizes leaders to capture state resources for the group they represent, at the expense of others, perpetuating identity conflict. Third-party interveners can either

20. "State building" refers primarily to strengthening the effectiveness of constitutional constraints and powers through their exercise in the arenas of government institutions such as the army and police, the judiciary and tax-collection agencies, and health and education systems. Fukuyama 2004. "Nation building" refers to forging individual-level identification with the nation, or the country. Hechter 2001; Mylonas 2012.

21. Lake 2016.

22. Brownlee 2007; Dobbins et al. 2003.

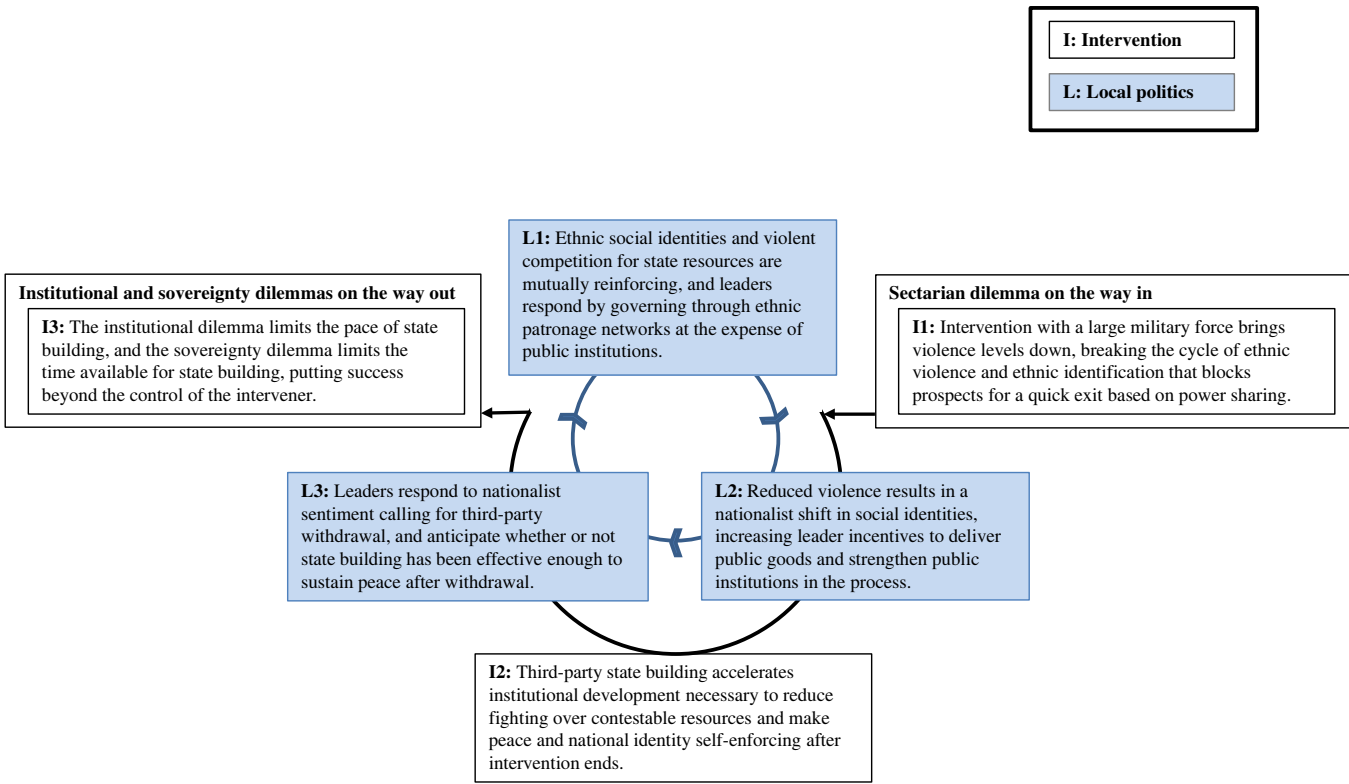


FIGURE 1. *The dilemmas of intervention in ethnic conflict*

risk a fragile peace and exit quickly, or intervene more heavily until leaders' incentives have changed. One channel for shifting leader incentives is when a third party brings a large military force (I1) to secure the population and bring fighting levels down to a level compatible with a unifying national identity (L2). As national identities become more salient, politicians are more likely to invest in public goods if they can reap political support from non-co-ethnics. By creating room for national identities to strengthen, intervention helps provide the foundation for self-enforcing peace.

But for this peace to last after foreign troops are withdrawn, local leaders must make such investments a credible commitment, aligning other stakeholders behind institutions and making it harder to steer resources to ethnic patronage networks in the future. This process takes time and gradually builds confidence in the population that leaders will respect constitutional rule, electoral results, and judicial authority. With each action that adheres to these rules, leaders may convince others to follow the rules and, more importantly, become invested in them, strengthening institutions over time. However, leaders are forward looking and may be cautious if they suspect that institutions will remain weak after the intervener withdraws, creating a self-fulfilling prophecy. To overcome this problem, the intervener can strengthen the state-building process with more resources, political facilitation, and technical assistance (I2). As leaders decide to invest in public goods, a virtuous cycle can gradually take hold as individuals observe leader behavior and leaders observe how the population shifts from ethnic to national identities.

This process is hindered by obstacles to the pace of institutional development and limits on the time available for state building before withdrawal (L3). Heavy-handed intervention to build institutions can crowd out leaders, which creates an institutional dilemma.

Institutional dilemma: A peace builder must help leaders deliver public goods to build institutional credibility, but taking too much control of that process limits local leaders' independent governing experience and does not allow them to develop reputations for competence. Then even cooperative leaders cannot confer credibility on institutions, which limits the state-building impact of intervention.²³

Under these conditions, leaders will be either ineffective in the task of institutional development or reluctant to cooperate with the occupier.²⁴ To limit crowding out, intervention must proceed slowly, but the sovereignty dilemma reduces the length

23. On how external state building can crowd out local leadership, see Call 2008; Chesterman 2004; Fukuyama 2005; Myerson 2009; Rubin 2008.

24. Evidence that foreign intervention can weaken the credibility of local leaders is emerging from experimental studies. For example, see Dietrich and Winters 2015 on the "branding" of foreign aid projects. Wimmer 2018 uses data from Afghanistan to show that government-funded projects are more than twice as effective as foreign-sponsored projects in increasing trust in government.

of time that foreign occupation will be acceptable as national identification becomes stronger.

Sovereignty dilemma: A third party's long-term commitment to security and state building makes it easier for local leaders to build institutions that will sustain cohesive nationalism, but also generates nationalist resistance to foreign occupation and incentives for leaders to expel the third party.

This dilemma is part of the contradiction of having external forces try to build legitimacy for the state. Nation building can strengthen state building when citizens identify with their country and they reward leaders for abiding by inclusive politics. The occupier can help by protecting the different communities while providing resources and organization to restructure the national army, reallocate public service positions, monitor corruption, and retrain judges and police officers—all types of institutional reforms that can induce citizens to identify with their nation. But how can an occupying force promote cohesive nationalism without also creating opposition to its very presence? Ironically, nation-building success can raise the pressure on local leaders to push the occupier out before the peace can be self-enforcing (I3).

If institutions are not strong enough by that point to sustain peace, leaders will not invest in state building, instead revitalizing their networks of ethnic support. This, in turn, can make ethnic identities more salient relative to national identities, and inter-group conflict and violence are likely to recur (back to L1).

Model Setup

To formalize this argument, we start with a baseline model that captures an internal conflict between two groups. Later, we amend the game to capture the effects of a third-party intervention. We model two groups, A and B , which we treat as unitary actors, and their leaders ℓ_A and ℓ_B , in a two-period game ($t = 1, 2$), in a society that produces a fixed quantity X of resources in each period. Total resources can be used to produce national public goods or ethnic (excludable) goods—or they can be used to support fighting. A share of total resources is available to the two leaders to allocate between public and ethnic goods, and the remainder is contestable through fighting. The amount that can be appropriated via fighting depends on the strength of state institutions that can protect resources from predation.²⁵

The level of institutional development (state capacity) in period t is denoted by $\kappa^t \in [0, 1]$, and it is taken as given since it depends on historical political development before period t . A minimum share X_0 of total resources is not appropriable, irrespective of the quality of state institutions (i.e., as long as there is a state and individuals in government, they retain some power to direct resources, even if there is no precedent

25. This approach follows Sambanis, Skaperdas, and Wolforth 2015 and is consistent with Besley and Persson 2010, where state capacity investments can generate growth and induce stability.

of its delivering public goods). Hence we write $X = X_0 + Y$, where Y represents the resources that can either be controlled by the state or are up for grabs through fighting, depending on the strength of institutions. Given state capacity κ^t , then, $X_0 + \kappa^t Y$ cannot be appropriated and is available to be allocated by the national government. For simplicity we assume that each leader ℓ_j controls half of the nonappropriable resources.²⁶

Leaders can allocate the resources under their control to produce either national public goods or ethnic goods.²⁷ The remaining $(1 - \kappa^t) Y$ is contested in ethnic fighting, and each group's share of this part of income is consumed exclusively by the group. The sequence of actions in each period t is as follows.

1. Each group $j \in \{A, B\}$ identifies ethnically ($I_j^t = 0$) or nationally ($I_j^t = 1$) and chooses how much it contributes to ethnic fighting, $f_j^t \in \mathbb{R}_+$.
2. Each leader ℓ_j chooses $c_j^t = 1$ or $c_j^t = 0$, which corresponds to whether they allocate the $\frac{X_0 + \kappa^t Y}{2}$ resources under their control to public goods or to ethnic patronage, respectively.

We let $\mathbf{f}^t = (f_A^t, f_B^t)$, $\mathbf{I}^t = (I_A^t, I_B^t)$, and $\mathbf{c}^t = (c_A^t, c_B^t)$.

Each group's single-period payoff consists of a material part and a psychological part. The material part is given by the sum of the amounts of public goods and ethnic goods the group gets. Group j 's share of the contested resources is determined by a contest function:

$$q_j(\mathbf{f}^t) = \begin{cases} \frac{1}{2} & \text{if } f_A^t = f_B^t = 0, \\ \frac{f_j^t}{f_A^t + f_B^t} & \text{otherwise.} \end{cases} \tag{1}$$

The amount of public goods produced by resources r is $\psi \kappa^t r$; that is, it depends on the state capacity, κ^t , and the technology of public goods production, $\psi > 0$. The material payoff for group j is

$$\pi_j(\mathbf{f}^t, \mathbf{c}^t; \kappa^t) = \frac{\psi \kappa^t (c_A^t + c_B^t)(X_0 + \kappa^t Y)}{2} + \frac{(1 - c_j^t)(X_0 + \kappa^t Y)}{2} + q_j(\mathbf{f}^t)(1 - \kappa^t)Y - f_j^t \tag{2}$$

26. The even split is the simplest way to capture the dynamics where both groups are a force in both fighting and politics; extensions could consider cases where one of the groups or leaders has more than half (or even all) of the state's control over resources, which in the current setup would generally lead them to erode the state on behalf of their ethnic group.

27. Leader allocations of resources are modeled as binary for simplicity and because the choices enter linearly into the leader utility functions. A potential extension is to consider diminishing marginal returns, since politicians might be able to glean benefit from nonbinary allocations while giving different justifications to different constituencies.

The first two terms of this expression represent the payoffs from resources controlled by the state. The first is from investment in public goods and reflects the contributions of both leaders and hence cross-ethnic support, and is mediated by the government’s institutions and technology for delivering public goods. The second reflects the payoff when a leader directs resources exclusively to their own ethnic group, bypassing the government machinery. The third and fourth terms represent the benefit and cost of fighting, respectively.

The groups’ payoffs also depend on identities. The payoff for group j in period t is

$$w_j(\mathbf{f}^t, \mathbf{I}^t, \mathbf{c}^t; \kappa^t) = \pi_j(\mathbf{f}^t, \mathbf{c}^t; \kappa^t) + I_j^t[\alpha - \beta(f_A^t + f_B^t)] \tag{3}$$

where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}_+$. The second term captures the influence of identity. The intuition is that there is a “belonging benefit” and a “heterogeneity cost” that enter group payoffs under national identification because each group feels some degree of attachment but also “social distance” from a nation that also includes the rival group (under ethnic identification, the second term in Equation (3) drops out, because groups care about only their own payoffs). These benefits and costs of identifying with the nation are partly shaped by exogenous factors that do not depend on current fighting levels (summarized by α), such as how proud people are of their nation (and therefore how likely they are to identify with it) or inter-ethnic hostility due to prior conflict. To the extent that national identity has been historically weak, α could be low (or even negative, though in our model this would exclude the possibility of a nationalist equilibrium). The salience of national identity also depends on the intensity of ethnic fighting: more fighting widens the perceived distance between ethnic groups, and therefore the distance from the nation grows. Fighting can also destroy national resources, reducing its standing/status, which would further reduce national identification. The term $\beta(f_A^t + f_B^t)$ therefore enters as a cost of identifying nationally, and β captures the sensitivity of national identification to fighting.²⁸

The leader’s payoffs for ℓ_j in the two periods are

$$u_j(\mathbf{f}^1, \mathbf{I}^1, \mathbf{c}^1; \kappa^1) = (1 - \mu_{-j}^1)w_j(\mathbf{f}^1, \mathbf{I}^1, \mathbf{c}^1; \kappa^1) + \mu_{-j}^1 w_{-j}(\mathbf{f}^1, \mathbf{I}^1, \mathbf{c}^1; \kappa^1) + \delta \left[u_j(\mathbf{f}^2, \mathbf{I}^2, \mathbf{c}^2; \kappa^2) \right] \tag{4}$$

$$u_j(\mathbf{f}^2, \mathbf{I}^2, \mathbf{c}^2; \kappa^2) = (1 - \mu_{-j}^2)w_j(\mathbf{f}^2, \mathbf{I}^2, \mathbf{c}^2; \kappa^2) + \mu_{-j}^2 w_{-j}(\mathbf{f}^2, \mathbf{I}^2, \mathbf{c}^2; \kappa^2) \tag{5}$$

The intuition here is that leaders’ payoffs depend on group payoffs, since leaders are modeled as caring only about shoring up political support. Thus, leaders’ payoffs will

28. This expression is based on a set of assumptions with broad support in the psychology literature: social identities influence behavior; the salience of identities changes over time; intergroup status comparisons affect social identification; and individuals prefer to identify with higher-status groups and with groups from which they feel less distant. Ethnic war lowers national identification by destroying resources (which lowers the nation’s status) and by widening social distance between ethnic groups. For a fuller discussion see Sambanis and Shayo 2013.

take into account their own group’s payoffs only if the rival group identifies ethnically; but if the population identifies nationally, then leaders could derive support from both groups, and they will take both groups’ welfare into account. Their utility function under national identification is therefore a weighted average of both groups’ welfare ($-j$ denotes j ’s rival group; i.e., $\{-j\} = \{A, B\} \setminus \{j\}$). Equation (4) gives group j ’s leader’s utility; the first term captures the utility from the leader’s own group’s welfare, and the second term captures the leader’s utility from the other group’s welfare when that other group identifies nationally. This second term reflects the utility leaders get from the cross-ethnic political relationship a unifying nationalism allows.

The weight assigned to the welfare of each group is captured by the parameter $\mu \in (0, \frac{1}{2})$, and we assume that leaders will always place greater weight on their own group. The third term in Equation (4) gives the leader’s utility in the second period, after occupation. Here δ is the discount factor reflecting how leaders weigh the post-occupation period. Unlike the public, which is not strategic, leaders are assumed to be farsighted and strategic, consistent with instrumentalist theories of ethnic conflict. But in our model strategic elites are embedded in a society where social identities can influence the public’s behavior, constraining leaders. This allows us to probe the limits of instrumentalist theories of ethnic politics.

Finally, state capacity, κ^t , varies over time, depending on the leaders’ choices. Let $\kappa^1 \in [0, 1]$ be exogenously given. For $t = 2$, $\kappa^t = \gamma\kappa^{t-1} + \frac{(1-\gamma)(c_A^{t-1} + c_B^{t-1})}{2}$, where $\gamma \in (0, 1)$. State capacity at the beginning of period t is a weighted average of capacity at the beginning of period $t - 1$ and the proportion of the national government resources that is used for public goods in period $t - 1$. Thus, as leaders invest in national institutions, they can contribute to state building incrementally. This modeling assumption captures the idea that institutions develop through use. Parameter γ captures how easy it is for leaders’ choices in the current period to change institutional capacity that is accumulated through past choices.

Baseline Case: No Intervention

A strategy for group j is a pair of functions $(f_j(\cdot), I_j(\cdot))$ where, for each given state capacity level $\kappa \in [0, 1]$, $f_j(\kappa)$ is group j ’s fighting level, and $I_j(\kappa) \in \{0, 1\}$ is group j ’s identity. We let $\mathbf{f}(\kappa) = (f_A(\kappa), f_B(\kappa))$ and $\mathbf{I}(\kappa) = (I_A(\kappa), I_B(\kappa))$. A strategy s for leader ℓ_j consists of a choice of $c_j^t(\kappa)$ for $t = 1$ and $t = 2$. For a given time period, we let $s(\kappa) = (s_A(\kappa), s_B(\kappa))$.

For the groups’ choices, we apply a variant of the concept of social identity equilibrium developed by Sambanis and Shayo.²⁹ We require that in equilibrium no group can be better off in terms of its current period utility by changing its fighting effort, given its identity and the other players’ strategies. Also, no group can be better off by

29. Sambanis and Shayo 2013.

changing its identity, given its fighting level and the other players' strategies. This solution concept requires consistency between identities and actions. Formally, we say that (f, I) is a social identity equilibrium given s if, for every $\kappa \in [0, 1]$ and every $j \in \{A, B\}$,

$$w_j(f_j(\kappa), f_{-j}(\kappa), I(\kappa), s(\kappa); \kappa) \geq w_j(f'_j, f_{-j}(\kappa), I(\kappa), s(\kappa); \kappa) \text{ for every } f'_j \in \mathbb{R}_+, \text{ and}$$

$$w_j(f(\kappa), I_j(\kappa), I_{-j}(\kappa), s(\kappa); \kappa) \geq w_j(f(\kappa), I'_j, I_{-j}(\kappa), s(\kappa); \kappa) \text{ for every } I'_j \in \{0, 1\}$$

In words: groups decide how much to fight given their social identities and, given the level of fighting, no group can be better off by changing its identity. The first condition is the standard Nash condition (groups choose their actions/fighting given their identities). The second condition, as in the Sambanis and Shayo model,³⁰ characterizes the process that determines social identification. We interpret this equilibrium as the state of the world in which the behavior of the groups and the psychological identities of their members are congruent and mutually reinforcing. The population does not maximize a farsighted payoff function while anticipating how they *might* identify under different fighting conditions. The groups are myopic with respect to fighting strategy (and choose it on the basis of their identity). They choose a social identity and a corresponding fighting strategy that reflects their environment (the level of fighting), and they are in equilibrium when that environment is consistent with these choices. Fighting strengthens ethnic preferences and lowers national status. In equilibrium, high fighting levels correspond to ethnic identification, and low fighting levels correspond to national identification.

Given that the two groups are identical, it is natural to focus on symmetric equilibria, that is, social identity equilibria with $f_A(\kappa) = f_B(\kappa)$ and $I_A(\kappa) = I_B(\kappa)$ for every $\kappa \in [0, 1]$. Let $f^e(\kappa) = \frac{(1-\kappa)Y}{4}$ and $f^n(\kappa) = \frac{(1-\kappa)Y}{4(1+\beta)}$. The following lemma characterizes symmetric social identity equilibria.³¹

Lemma 1: For every profile of strategies for the leaders, s, the following statement is true. A profile of the groups' strategies (f, I) is a symmetric social identity equilibrium given s if and only if:

- 1 If $I(\kappa) = (0, 0)$, then $f(\kappa) = (f^e(\kappa), f^e(\kappa))$ when $\kappa \leq 1 - \frac{2\alpha}{\beta Y}$.
- 2 If $I(\kappa) = (1, 1)$, then $f(\kappa) = (f^n(\kappa), f^n(\kappa))$ when $\kappa \geq 1 - \frac{2\alpha(1+\beta)}{\beta Y}$.

In general, if state capacity is high, then the groups identify nationally because high state capacity implies less contestable resources, which reduces fighting,

30. Sambanis and Shayo 2013, 10.

31. In the online supplement we show that the symmetric equilibrium is unique for both the ethnic and national-identity equilibrium cases, with the additional stipulation that where state capacity allows ethnic and nationalist equilibria, both groups identify in the same way.

which in turn reduces the salience of ethnic identification. In addition, there can be multiple equilibria because the conditions on κ are not mutually exclusive. In every equilibrium, when $\kappa > 1 - \frac{2\alpha}{\beta Y}$, the groups' identities are national, and when $\kappa < 1 - \frac{2\alpha(1+\beta)}{\beta Y}$, the groups identify ethnically. But when $1 - \frac{2\alpha(1+\beta)}{\beta Y} \leq \kappa \leq 1 - \frac{2\alpha}{\beta Y}$, the identities may be ethnic or national depending on which equilibrium we choose.

Regarding the issue of multiple equilibria, in our subsequent analysis, we choose the equilibrium with the largest set of state capacity in which groups identify nationally.³² Formally, let $\hat{\kappa}(\alpha, \beta) = 1 - \frac{2\alpha(1+\beta)}{\beta Y}$. Define (f^*, I^*) by the following: for every $j \in \{A, B\}$,

$$(f_j^*(\kappa), I_j^*(\kappa)) = \begin{cases} (f^e(\kappa), 0) & \text{if } \kappa < \hat{\kappa}(\alpha, \beta), \\ (f^n(\kappa), 1) & \text{if } \kappa \geq \hat{\kappa}(\alpha, \beta) \end{cases} \tag{6}$$

From Lemma 1, it is obvious that (f^*, I^*) is a symmetric social identity equilibrium. From now on, we assume that groups play (f^*, I^*) .

The higher κ is, the lower are the stakes of conflict, and therefore the relationship between fighting levels and the strength of state institutions is a decreasing function. Since fighting levels are interdependent with social identity, there is a stepwise decrease in violence separating the case where the population identifies ethnically (high fighting) from the case where the population identifies nationally (low fighting). This result is illustrated by the dashed line in Figure 2 (where we also anticipate the result of the impact of an intervention force; see next section).³³

The critical determinant for national identification is institutional strength, which in turn depends on the choices leaders make. Institutions get incrementally stronger the more leaders use them. Thus leaders in a country with initial levels of state capacity below $\hat{\kappa}$ can help break the conflict trap by investing in institutions, up to the point where the population identifies nationally, making postwar peace sustainable. But why would leaders choose this path? We want to capture the possibility that leaders act strategically and can contribute to institutional development or deterioration, which would affect their own future payoffs. The question is then whether they can create a situation where the rewards for public goods delivery are higher than for ethnic patronage in the future and they are farsighted enough to value this

32. It is natural to select an equilibrium in which there is a cutpoint of state capacity such that the population identify ethnically below the cutpoint and nationally above the cutpoint. Among a continuum of cutpoint equilibria, we choose the equilibrium with the lowest cutpoint. As long as we choose one of the cutpoint equilibria, our result would not change qualitatively.

33. In the figure, $Y=4, \alpha=0.5, \beta=1, \hat{\kappa} = 0.5, \kappa^1 = \hat{\kappa}' = \frac{1}{4}, f_a = \frac{3}{4}, f_a' = \frac{1}{4}$, and $f_c = \hat{f}_c = \frac{1-\sqrt{3}}{4}$, with the value for \hat{f}_c given by Equation (8) in the online supplement, ensuring that κ^1 becomes consistent with a national identity equilibrium under intervention. Note that the dashed lines (nonintervention) are given by $f_a = \frac{(1-\kappa)Y}{4}$ for the range of $\kappa < \hat{\kappa}$ where the population identifies ethnically and $f_a = \frac{(1-\kappa)Y}{4(1+\beta)}$ for $\kappa \geq \hat{\kappa}$ where the population identifies nationally; for intervention, the solid lines are given by Equations (5) and (7), respectively, in the online supplement. These equations are approximately linear over the range shown (with a slight downward curve that becomes more pronounced below $f_a=0$, eventually joining at $f_a = \frac{-f_c}{2} = -0.34$, but that is outside the relevant domain).

future and to bear the short-term cost of acting counter to the preference of their own base. This kind of leadership in local politics becomes the critical lever for breaking the conflict trap. The two-period game allows us to capture this intertemporal trade-off in the simplest way possible: if κ in period 2 is above $\hat{\kappa}$, it induces a nationalist shift in the public, and peace is self-enforcing among leaders, institutions, and the two groups. We focus on cases where $\frac{1-\mu}{\psi} \leq \hat{\kappa} \leq \frac{1}{\psi}$, since that is the range of theoretical interest, where leader choices depend on social identification of the population groups (below the lower threshold, leaders deliver only for their group in the one-shot game of period 2, whereas above the higher threshold they deliver public goods, in both cases regardless of social identities of the groups; see online supplement for details).

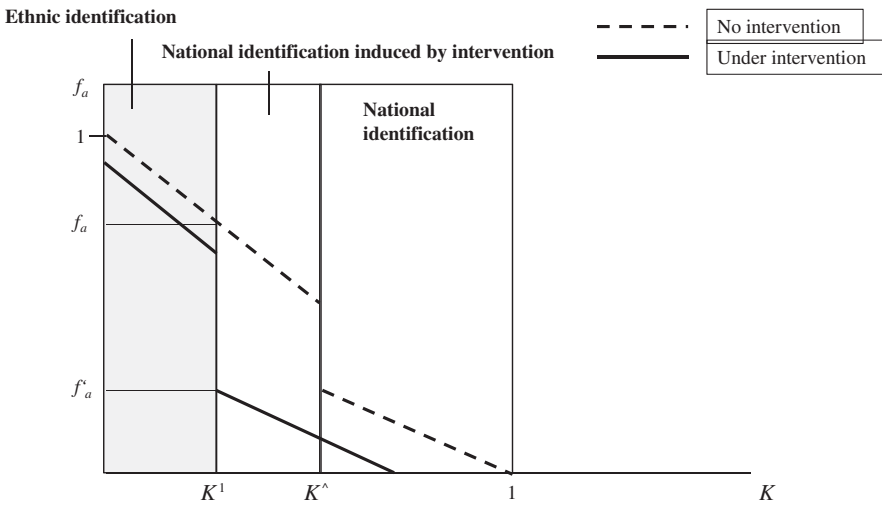


FIGURE 2. Social identity shift to nationalism induced by intervention

The requirement that $\kappa^2 > \hat{\kappa}$ implies a threshold for γ , which is the determinant of institutional strength in period 2 for given leader choices in period 1. We focus on γ because it determines the impact of leader choices over time, anticipating that this is the main consideration for the success of interventions that are limited by time, which our theory will address. Whether $\kappa^2 > \hat{\kappa}$ will also depend on the other parameters of the model that influence the likelihood of different leader choices, which we also indicate. The γ threshold depends on whether just one leader chooses public goods in period 1 and has enough influence to unilaterally strengthen institutions above $\hat{\kappa}$, or if both choose public goods. These thresholds give the following results (proofs in the online supplement).

Proposition 1: For a given initial institutional capacity $\kappa^{t=1}$ and parameters defining the depth of the ethnic conflict (α, β, Y), there is a range of institutional resistance

to change, $\bar{\gamma} < \gamma < \bar{\bar{\gamma}}$, in which $\kappa^2 > \hat{\kappa}$ if and only if both leaders choose public goods and where, under certain conditions of (α, β, Y) , farsighted leaders will pursue self-enforcing peace.

Proposition 1 means that self-enforcing peace is possible, but (as the online supplement shows) only when leaders are farsighted and nationalism has a high payoff, in that second period, will they invest in institutions and public goods. Leaders anticipate the payoff from nationalism that public goods provision will induce in the second period (i.e., potential support from non-co-ethnics).³⁴

Proposition 2: For a given initial institutional capacity $\kappa^{I=1}$ and parameters defining the depth of the ethnic conflict (α, β, Y) , there is a threshold for institutional resistance to change, $\bar{\bar{\gamma}}$, with $\bar{\bar{\gamma}} < \bar{\gamma}$, below which one leader can unilaterally induce $\kappa^2 > \hat{\kappa}$ and where, under certain conditions of (α, β, Y) , farsighted leaders can reach self-enforcing peace.

The values of $\bar{\gamma}$ and $\bar{\bar{\gamma}}$, and therefore the likelihood that a leader can achieve peace for a given κ^1 generally increase with α (the exogenous determinants of national identification) and decrease with β (the sensitivity of national identification to ethnic fighting). A higher α means that a cohesive nationalism can emerge at a lower level of institutional strength (i.e., $\hat{\kappa}$) because of a shared identity that persists unchanged in conflict; and as β decreases it means that this nationalism is more resilient to losses of national resources and the increased saliency of ethnicity due to ethnic fighting (i.e., the erosion of nationalism is less than with high β). Similarly, the value of the γ thresholds generally decreases (more state building is needed) with Y , since contestable resources make greater institutional strength necessary for stable nationalism. Furthermore, $\gamma < \bar{\bar{\gamma}}$ does not guarantee that at least one leader will choose to invest in public goods as self-enforcing peace requires. Under the conditions of Proposition 2, where one leader can move institutional strength above $\hat{\kappa}$, leaders will only do so if they are farsighted enough and the payoffs for nationalism in period 2 are large enough.³⁵

34. As the online supplement shows, leaders will have a dominant strategy to invest in public goods if they are farsighted under a narrow range of κ^1 that is either very low or close to the $\hat{\kappa}$ threshold, but more generally will face a coordination game, and they have reason to pursue coordination: it is better for each leader to take this path if the other one also does (i.e., it is not a prisoner's dilemma, since matching the other leader is precisely what is needed to get above $\hat{\kappa}$ and the higher payoff that can offer).

35. As the online supplement shows, these conditions again create a dominant strategy among reduced form options or a coordination game for both leaders to invest in public goods. The potential for one leader to unilaterally induce a nationalist shift between periods 1 and 2 also creates the possibility of a game of chicken, where each leader wants to reap ethnic rewards in period 1 by delivering for their ethnic group while benefiting from nationalism in the second period as a result of the other leader's period 1 public goods investment; while the possibility exists, the online supplement shows that it is only under very narrow conditions, with high X_0 , high γ (close to $\bar{\bar{\gamma}}$), and low δ .

Taken together, Propositions 1 and 2 reflect how difficult it is to arrive at self-enforcing peace when leaders have reached their position of power in the context of ethnic identity and fighting due to low α and high β and Y relative to κ , even when they are acting instrumentally (except, as the online supplement shows, when the efficiency of converting resources into public goods, ψ , is so high that leaders' choices don't depend on social identities, which is of less theoretical interest). Farsightedness is a critical condition, but if γ is above $\bar{\gamma}$ then even farsighted leaders will have little interest in acting for a more unified public good because state building to make national identity sustainable and public goods rewarding is too difficult. Can intervention make leaders' investments in public goods more likely and more effective, increasing the prospects for self-enforcing peace?

Intervention

In the previous section we showed that, when left to their own, leaders will not have incentives to invest in public goods and build national institutions when fighting and ethnic identities are reinforcing unless they become more farsighted or nationalism exogenously becomes more attractive. Can foreign intervention help resolve this problem? Here, we amend the game to capture the effects of foreign intervention. In doing so, we abstract from the incentives of the interveners and instead consider the conditions an intervention would have to meet to increase the chances of lasting peace. A clarification about timing is in order. In the informal discussion of the theory, we highlight the concern over how long the occupier can stay and the possibility that they will be pushed out prematurely. In the model, since intervention always lasts one period, this corresponds to a question of whether the occupier's investments can move state capacity enough in the first period to make a difference.

We envision intervention as taking two forms. In the first, the third party commits forces in a way that diminishes the gains from ethnic conflict. In the second, it invests resources in reconstruction and technical assistance, which increases the state's ability to provide public goods. We consider each of these in turn.

If the third party commits the forces necessary to break the reinforcing equilibrium of high fighting levels and ethnic social identity, this can affect leaders' actions via the social identification mechanism.³⁶ In terms of the model, the contest function for the first period (corresponding to occupation) becomes $q_j(f^t) = \frac{f_j^t}{f_a + f_b + f_c}$, where f_c is the third party's fighting resources. Referring back to Figure 2, fighting by the intervener diminishes the fighting payoffs for A and B , leading to lower levels of f_a at any given level of institutional strength, κ . We assume that third-party forces do not contribute to ethnic polarization, which means that both the slope and the point of discontinuity, $\hat{\kappa}$, in Figure 2 are lower than in the case without intervention.

36. This idea is similar to the role of "bottom-up" population security provision to build government legitimacy in counterinsurgency doctrine; we go further to focus on the social identity dimension.

There will be a minimum $\widehat{f}_c(\kappa)$ such that given κ , intervention with this level of third-party fighting effort will reduce conflict enough to lead to national identification. Specifically, this will be where $f_a = \frac{\alpha}{2\beta}$, which is where the psychological component of national identity turns positive in the group payoffs (and is therefore a social identity equilibrium as long as fighting levels are optimal under conditions of national identity, as given by Equation (7) in the online supplement and the lower solid line in Figure 2). A third-party force can therefore reduce ethnic fighting to the point where the current institutional strength κ^1 is a national identity equilibrium, and group A's fighting level drops from f_a to f'_a in the figure.

Since we focus on interventions that aim to produce an inclusive political system and incentivize nation building, it is natural to assume that the occupier does not extract resources by fighting.³⁷ Increasing third-party forces beyond \widehat{f}_c reduces fighting levels further in our model, but will not affect leader choices. A commitment of forces below this critical threshold means that fighting levels are not reduced enough to shift social identities away from sectarianism, and the nation-building path to peace fails. Whether intervention will achieve this goal depends on the intervener's incentives, which we do not model. We focus on cases where the intervention is sufficiently large to incentivize the two groups to choose national identity in the first period.

If the occupier commits forces of at least \widehat{f}_c , then forward-looking leaders will naturally consider whether the resulting nationalist environment alters their own incentives and whether such conditions will last. In the first period of our two-period game, leaders encounter a public that identifies with the nation in a low-fighting environment. These conditions make it more likely that leaders will choose to deliver public goods in response, reaping the reward of non-co-ethnic support, as specified in Propositions 3 and 4.

Proposition 3: For a given initial institutional capacity ($\kappa^{t=1}$), parameters defining the depth of the ethnic conflict (α, β, Y), and an intervention force that brings fighting levels to a nationalist equilibrium in period 1, and where $\gamma < \bar{\gamma}$, leaders are more likely to invest in public goods and reach self-enforcing peace than under the same conditions without intervention (i.e., compared to Proposition 1).

Proposition 4: For a given initial institutional capacity ($\kappa^{t=1}$), parameters defining the depth of the ethnic conflict (α, β, Y), and an intervention force that brings fighting levels to a nationalist equilibrium in period 1, and where $\gamma < \bar{\gamma}$, each leader is more

37. The share of resources under the occupier's control is part of the endowment that cannot be appropriated via fighting by the parties. We assume an equitable distribution of those resources, so as to accommodate arguments that nationalist opposition to alien rule will be mitigated if there is an equitable distribution of public goods. Hechter 2013; Lawrence 2013. Thus our model does not apply to extractive occupations or colonialism.

likely to invest in public goods and reach self-enforcing peace than under the same conditions without intervention (i.e., compared to Proposition 2).

The γ thresholds do not change from the nonintervention case because the same amount of institution building needs to happen in period 1 to achieve self-enforcing peace in period 2 after the intervening force is gone. Nonetheless, even without any changes to how effective leaders are at state building (i.e., $c^1 = (1, 1)$ still has the same impact on κ^2), Propositions 3 and 4 indicate that intervention that is limited to security provision could be enough to overcome the sectarian dilemma and create self-enforcing peace by making public goods investment by leaders more *likely*. Whether this strategy can work depends on the parameter μ , which captures the value of cross-ethnic support that leaders get when they invest in public goods, and this is possible when the population identifies nationally (see Equation (4)). If this value is high, it becomes more likely that forceful intervention will shift the population to national identification; if it is low, then intervention is unlikely to have such a pacifying effect.

This nation-building channel can work only if leaders on their own can do the state building necessary to make peace sustainable after withdrawal, which is when $\kappa^2 > \hat{\kappa}$. What if that is not the case, in other words if γ is above even $\bar{\gamma}$? In this unfavorable case, the best that can be achieved is a temporary peace that will fall apart once the intervener departs. This temporary peace can obtain: if leaders are not too farsighted in what they value they may choose to deliver public goods to their nationalist population, even knowing that it will not be enough to avoid a return to ethnic behavior after withdrawal. Such a possibility leads to the following proposition.

Proposition 5: For a given initial institutional capacity ($\kappa^{t=1}$), parameters defining the depth of the ethnic conflict (α, β, Y), and an intervention force that brings fighting levels to a nationalist equilibrium in period 1, and where $\gamma > \bar{\gamma}$, both leaders can have a dominant strategy in the reduced form of the game to choose public goods in period 1 even though the outcome will be ethnic conflict after withdrawal.

Proposition 5 implies a kind of “double game” where leaders present a nationalist face to the public and the intervener alike while expecting an ethnic turn after withdrawal, a result we consider more closely in the case of Iraq.

The second form of intervention we envisage involves moving beyond counterinsurgency and investing resources and technical assistance to help leaders build state institutions. We allow for the possibility that such a strategy can make nationalist choices by leaders in period 1 not only more *likely* but more *effective*, and thus be enough to sustain peace in period 2 after withdrawal. For the former we allow for an expanded contribution to public goods delivery by the third party and rewrite the group payoff as

$$\pi_j(f^t, c^t; \kappa^t) = \frac{\psi \kappa^t (c_A^t + c_B^t + c_C^t(\chi))(X_0 + \kappa^t Y)}{2} + \frac{(1 - c_j^t)(X_0 + \kappa^t Y)}{2} + q_j(f^t)(1 - \kappa^t)Y - f_j^t \tag{7}$$

Here $c_C \in \mathbb{R}_+$ is the impact of (in principle limitless) third-party resources, which we denote χ , on the country’s public goods delivery. In principle this allows an arbitrarily large incentive for leaders to choose public goods in period 1, if we assume that c_C monotonically increases with χ (and with $c_C(0) = 0$). In other words, if an intervention is either heavy-handed enough (through trusteeship in the extreme case) or operates with a light footprint (in terms of f_C) but a consent-based mandate that enables high χ nonetheless, it can control the direction of resources toward public goods.

However, this will do no good for sustainable peace if it does not strengthen institutions in the process. Here the institutional dilemma arises: the virtuous circle of the population’s unifying national identity, leader choices to deliver public goods, and the strengthening of institutions as a result will not emerge if the public believes that leaders are working through institutions only because of incentives or threats from the occupier. We therefore consider the impact of χ on the equation that relates institutional strength in one period to institutional strength in the next and rewrite it as

$$\kappa^{t+1} = \gamma \kappa^t + (1 - \gamma) \frac{c_A + c_B + c_C(\chi) - \omega(\chi)}{2} \tag{8}$$

This equation is modified from the basic game through the introduction of two terms: $c_C(\chi)$, which (as before) captures the positive effect of the intervener’s investment in additional resources and technical assistance when they have local leaders to work with; and ω , a new parameter that captures the negative effects of foreign assistance on domestic institutional development. $c_C(\chi)$ enters the equations for both leader incentives and institutional development in a similarly positive way, reflective of the way interventionists generally frame the possibilities for third-party state building. But the third party acting in the place of legitimate local leaders could diminish the impact of the third-party contribution. Parameter ω reflects the severity of the institutional dilemma and requires that $\omega(0) = 0$ and $\frac{\partial \omega}{\partial \chi} > 0$. That is, the heavier the intervention in local governing processes, the weaker is the role of local leaders and the more likely it is that they will be seen as puppets. Taken together, these adjustments to the model reflect that intervention can boost the potential for local leaders to effect change, but the impact is muted to the extent that the occupier crowds them out.

To maximize impact, the occupier must balance these two considerations, which occurs where $\frac{d\kappa^{t+1}}{d\chi} = 0$; for the case of interest, where $c^{t=1} = (1, 1)$, this obviously occurs where $c_C(\chi) - \omega(\chi)$ is maximized, due to our construction of Equation (8) in reduced form. This most clearly shows the third party’s intervention as a trade-off.

We assume that leaders and third-party partners can manage public communication and ensure that leaders get credit for delivering public goods as long as χ is low; the institutional dilemma kicks in as χ gets large and it becomes obvious that leaders are dependent on third-party resources for public goods delivery. In other words, we require that $c'_C(\chi = 0) > \omega'(\chi = 0)$ and $c''_C(\chi = 0) < \omega''(\chi = 0)$. This will generally give a solution $\chi = \chi^*$ that is the level of third-party resources that maximizes κ^{T+1} . These are the conditions for third-party-led state building. Once χ reaches χ^* , any additional push from the third party will have a detrimental effect, as additional resources are offset by the crowding-out effect. These results are captured in Proposition 6 (proof in the online supplement).

Proposition 6: An intervention force can increase the threshold of institutional resistance to change $\bar{\gamma}$ and $\bar{\gamma}$ in Propositions 3 and 4, respectively, by increasing state-building resources, χ , to a maximum level χ^ , which is determined by the severity of the institutional dilemma, given by $c_C(\chi) - \omega(\chi)$.*

Several factors, such as the occupier's identity, technical capacity, and perceived impartiality, and the heavy-handedness of the intervention (f_c), determine the slope of $\omega(\chi)$ and thus χ^* and κ^{T+1} and the prospects for success. We explore these empirical possibilities in the next section, in particular the differential rate of success in light versus heavy-handed and unilateral versus multilateral interventions, where consent-based multilateral interventions can mitigate the institutional dilemma (i.e., flattening the slope of $\omega(\chi)$).

If the occupier optimizes state-building resources given the trade-off defined by the institutional dilemma, the question then becomes whether the occupation period can last long enough during rising national identification. The sovereignty dilemma then arises from the assumption that when a population shifts to national identification and violence decreases, foreign occupation is no longer feasible for political leaders riding a nationalist sentiment.³⁸ Social identification implies caring about the relative status of one's group and preferring outcomes that increase the group's status. Thus to identify nationally means to oppose anything that diminishes the country's status, as occupation readily can. Just as foreign intervention begins to succeed in inducing leaders to develop national institutions, that very success limits the options for continued foreign engagement.

The structure of the sovereignty dilemma in our theory parallels the institutional dilemma, in that a greater reward for choosing public goods in a given period carries with it a trade-off in terms of the impact on institutional development. In the case of the sovereignty dilemma, a sense of nationalism exogenous to ethnic

38. If the population still identifies ethnically, the institutional dilemma still applies and slows the pace of third-party state building but can go on for longer due to the absence of a strong sovereignty dilemma, as we describe in the empirical section for the case of Afghanistan.

fighting, captured by α , plays that role. The positive impact of this nationalism is already captured in Equation (3). To capture the trade-off we now require that ω is also an increasing function of α , with

$$\kappa^{t+1} = \gamma\kappa^t + (1 - \gamma) \frac{c_A + c_B + c_C(\chi) - \omega(\chi, \alpha)}{2} \tag{9}$$

In other words, for a given $\chi = \chi^*$, the more severe is the sovereignty dilemma, the higher is $\omega(\chi^*, \alpha)$ and the smaller is the total impact of third-party intervention on institutional strength in period 2. Since α is fixed for a given intervention rather than a choice of the intervener, there is no optimization question, and we assume that $\omega'(\chi)$ is independent of α , such that the impact of the sovereignty dilemma on $c_C(\chi) - \omega(\chi, \alpha)$, and therefore the γ thresholds in Proposition 5, are fixed, further reducing the space for successful peace building by that incremental amount. The severity of the sovereignty dilemma (the value of $\omega(\chi, \alpha) - \omega(\chi, 0)$) will depend only on the identity of the intervener and is triggered only when the heavy-handedness of f_c is enough to induce a nationalist shift in the population.

The model of how interventions affect the local politics of ethnic conflict suggests a two-part framing for debates over intervention: (1) whether an intervention force can commit and calibrate resources to reach \hat{f}_c and χ^* and thus overcome the sectarian dilemma and maximize the chances of success; and (2) whether, given these commitments, local politics will indeed lead to enough state building to succeed before the occupier is pushed out. In the empirical section, we use this framework to integrate the main debates over intervention in a simple way and illustrate the empirical implications.

We should note that a scope condition for our theory is our assumption that foreign rule is antithetical to nationalism. While this is a consensus view in the literature, a few studies argue that “alien rule” can be acceptable as long as the ruler provides public goods and distributes them equitably.³⁹ Our theory can capture such cases, which are likely a minority, via the parameters designed to describe the sovereignty dilemma. If the country’s history is such that national identification is not antithetical to being colonized, or if the identity of third-party interveners is such that occupation does not elicit an anti-imperial reaction, then $\omega(\alpha)$ will be low, and therefore κ^{t+1} will be higher. However, in the presence of a sovereignty dilemma, there are inherent limits to what the third party can achieve, even if intervention is well managed: the inducement of national identification needed for success pushes leaders to expel the occupier, especially after violence has dropped and security has been restored. Thus we do not see a linear relationship between collective goods provision and resistance to occupation. That relationship is contingent on how the social identities of the population interact with leader incentives at different levels of state capacity and for different interveners.

39. See Hechter 2013; Lawrence 2013. Indeed, consistent with these studies, we model an occupier who is impartial, and improved state capacity benefits both groups evenly. We see this type of public provision as a necessary (but not sufficient) condition for the success of foreign occupation.

Empirical Applications

Our theory shows that sustainable peace depends on a reinforcing equilibrium between social identities and leader choices, not just power sharing among elites. Getting this virtuous cycle going is hard and requires the commitment of resources to bring fighting down, and effective collaboration with local leaders to build institutions. Even then, the two exit dilemmas (institutional and sovereignty) create challenges for nation-building success.

These dynamics create two tasks for any intervention. The intervener must first mobilize resources to the right level to overcome the sectarian dilemma. In the model’s terms, this implies providing enough resources to reach \hat{f}_c and χ^* , that is, to shift to a nationalist equilibrium and accelerate state building as much as possible. Such a heavy-handed intervention might not always be necessary. Light-footprint interventions—which are more limited in coercion, duration, and investment of resources for state building—may be preferable in some circumstances due to lower cost, cross-border sanctuary, or local support for the fighting groups, which would limit the effectiveness of an intervention force, or in conflicts where third parties can build institutions with the help of local leaders even while the population identifies ethnically. But where the sectarian dilemma is strong and local leaders owe their power to ethnic constituencies, and where building self-enforcing peace requires investments in state building, a heavier footprint will likely be needed to change leader incentives.

Second, before committing to a heavy footprint, the intervener must also weigh the risk that an intrusive intervention could undermine local leaders and be perceived as a threat by the population as it begins to identify nationally. These sources of opposition could turn leaders against third-party-led state building. In the model’s terms this means assessing the exit dilemma parameter $\omega(\chi^*, \alpha)$, which determines the potential for state building during occupation with a unified population. Military and state-building investments must be done in a way that does not alienate local leaders, and even then success may be out of reach.

This tension is captured in the two dimensions of Table 1, which presents a matrix of peace-building possibilities based on the two considerations facing the third party: the level (intrusiveness) of third-party resources (relative to the \hat{f}_c threshold) and the perceived threat to leaders and sovereignty of this intervention, which depends on

TABLE 1. *Intervention diagnostic*

		<i>Intervener identity / perceived threat to sovereignty</i>	
		<i>High threat</i>	<i>Low threat</i>
Level of intrusion	Light footprint	(1) Least success	(3) More success
	Heavy footprint	(2) Less success	(4) Most success

$\omega(\chi, \alpha)$. Holding the factors determining the severity of ethnic hostilities (α, β, Y) constant—and assuming an intervention that is not resource-constrained—Table 1 presents a simple diagnostic framework for the likely outcome of different types of intervention in civil wars.

To make things concrete, we apply our theory to discuss the US intervention in Iraq in 2003 to 2011, and we explain why the diagnostic predicts the ordering of the likelihood of peace-building success that is depicted in Table 1. We then use the diagnostic tool to discuss humanitarian intervention, foreign-imposed regime change, peacekeeping, and neo-trusteeship.

Peace-Building Failure in Iraq

The period from 2003 to 2006 in Iraq was dominated by debates about whether the United States had committed enough troops to provide security and whether it could achieve a representative balance of the various ethno-sectarian communities in Iraq's new "unity" government. Resolving the sectarian dilemma through representative power sharing to enable transition to a new Iraqi government was in the forefront of these debates.⁴⁰ However, it did not work, and sectarian fighting worsened. As the conflict transformed from an insurgency against the US occupation, driven mainly by the Sunni factions ousted from power, to a sectarian civil war, fighting and sectarian identities became mutually reinforcing, and US intelligence assessed that violence had become self-sustaining.⁴¹

The sectarian dilemma was evident in full force, with many voices in Congress and in the Iraq Study Group calling for a draw-down based on a transition to Iraqi security and political responsibility, accepting that violence would persist—while the administration instead opted for the surge strategy, to try to secure the population and strengthen the state and legislation to deliver public goods. While a number of factors affected the subsequent drop in violence,⁴² the Iraq case illustrates how the sectarian dilemma naturally raises the question of whether to try to transform the conflict, which has direct implications for how to deploy third-party resources. This is precisely the first dimension of our intervention diagnostic.

The Iraq case also shows what we mean by the second dimension in Table 1, which amounts to how much state building an intervention can achieve without undermining itself through a perceived threat to sovereignty. The institutional and sovereignty dilemmas limit the duration and pace of state-building intervention if the perception of sovereignty threat is high. US policymakers in Iraq confronted and debated this dilemma explicitly. As President Bush explained to a reporter, "There is a delicate

40. See the interview with Paul Bremer in "Did the US Occupation Create ISIL?" *Al Jazeera TV*, 1 December 2015, retrieved from <<https://www.aljazeera.com/program/head-to-head/2015/12/4/did-the-us-occupation-create-isil>>.

41. Woodward 2008.

42. Biddle, Friedman, and Shapiro 2012.

relationship between self-sufficiency on the Iraqis' part, and US presence."⁴³ Senator Collins asked General Petraeus during confirmation hearings whether the new strategy in 2006 went against T.E. Lawrence's dictum that it is "better they do it imperfectly with their own hands than you do it perfectly with your own," a position General Casey, and Petraeus himself, had previously emphasized.⁴⁴ General Petraeus responded:

What you described really has been truly an intellectual tension, frankly, about the mission in Iraq all along ... There are times when they start to wobble and the question is when do you move back in and provide assistance. In the wake of the bombing of the Samarra mosque and the violence that escalated throughout the latter part of 2006, I think we have arrived at a point where in fact we do need to help them a bit more in providing security in particular.⁴⁵

In the event, violence decreased from over 2,000 "ethno-sectarian deaths" per month to under 200 per month within a year.⁴⁶ By 2009 most American policy analysts agreed that the surge had created "bottom-up" demand for reconciliation in the population, but not yet "top-down" political decisions that could sustain peace.⁴⁷ Consistent with our model, the 2009 provincial elections were marked by parties competing to claim the nationalist mantle,⁴⁸ with Maliki's State of Law Party doing particularly well, at the expense of the Islamic Supreme Council of Iraq (ISCI, formerly SCIRI), which had been ascendant in the earlier period of sectarian balancing.

This account does not turn on whether Maliki himself was sectarian but on the model's assumption that as a leader he could benefit by responding to the nationalist turn in the population. He became prime minister at a time when the United States wanted leaders they could manage and who could have the support of their sect on the one hand and reach compromises with other sectarian leaders on the other. As a low-profile Dawa Party leader who didn't threaten either of the two poles of Shia power (the Sadrist and SCIRI), Maliki fit these criteria.⁴⁹ With the shift to the surge strategy, some US policymakers were concerned that a weak Shia leader from an Islamist party might not be right for a more aggressive stand against sectarianism.⁵⁰ Nonetheless, when violence dropped steeply and the population turned

43. Woodward 2008.

44. Smith 2016, 497.

45. Testimony to the Senate Committee on Armed Services, 23 January 2007, available at <http://fas.org/irp/congress/2007_hr/sasc.pdf>.

46. Department of Defense, "Measuring Stability and Security in Iraq," 26 September 2008, available at <<https://www.health.mil/Reference-Center/Reports/2008/09/26/Measuring-Stability-and-Security-in-Iraq>>.

47. Biddle, O'Hanlon, and Pollack 2008; Ricks 2009.

48. Visser 2011.

49. Ali Khedery, "Why We Stuck with Maliki and Lost Iraq," *Washington Post*, 3 July 2014.

50. An early indication prior to the shift in US strategy was that Maliki gave General Casey a "don't touch" list of political targets, starting with Sadr. Woodward's detailed account of the initial 22 July 2006, White House strategy review also quotes Hadley: "Are we convinced that

nationalist, Maliki benefited by casting his bloc, the State of Law Party, as a source of order. This contrasted with religious parties like ISCI, which people blamed for the sectarian violence, and Maliki's bloc made huge gains in the 2009 provincial elections.

The shift in social identities toward Iraqi nationalism was possible even though the concept itself has been fiercely contested because it has also been highly relevant to cohesion in the face of colonialism. National, ethnic, religious, tribal, and class identities have all shaped Iraqi history. Islamic sectarianism, born in Karbala in the seventh century, has periodically been important to Iraq's politics, as at the time of Iraq's modern founding and under Saddam's rule. So too have tribal and class identities been equally important. Yet the national identity was clearly "available" in the sense that Iraqi nationalism did not have to be fabricated from scratch. Rather, it had its roots in the 1920 Revolution against the British, the anticolonial appeal of pan-Arabism, and the 1958 Revolution, which was a nationalist opposition to a foreign-tainted monarchy.⁵¹

Throughout most of its history since World War I, including the Baathist period after 1968, Iraq was governed by leaders with narrow groups of trusted supporters in the military, clinging to power through the patronage apparatus offered by the oil state and both responding to and shaping ethno-sectarian and nationalist identities.⁵² Nationalism took different and contested forms. For example, the Iran–Iraq war created a lasting rift between Shia "exiles"—especially those who fought on Iran's side and remained there after the war—and "nationalists" who deepened their attachment to the state through rejection of any foreign interference, a view often expressed by Mohammad Sadeq al-Sadr after the Iran–Iraq war. At the same time, while the pan-Arabist orientation of an earlier variant of Iraqi nationalism was committed to Iraq's unity and status as a regional power, it nonetheless preserved a dominant role for Sunnis,⁵³ and the 1991 intifada after the Gulf War revealed deep Shia resentment of Sunni-dominated rule. Yet, while it could not resolve differing attachments to pan-Arabism, nationalism in Iraq offered an alternative to sectarian commitments and always evoked anti-imperialist sentiment.⁵⁴

The success of the surge allowed Iraqi leaders to get credit for nationalist choices, as Maliki largely did in his confrontation with Sadr's militias in Karbala in 2007 and then Basra in March 2008. US forces pulled back and deliberately managed the institutional dilemma by ensuring that Iraqi leaders got credit for reconstruction progress.⁵⁵ At the same time, this very dynamic of re-emerging Iraqi nationalism and

Shia leaders in Baghdad are serious about reining in the JAM and Shia death squads?" To which Casey responded, "No." Woodward 2008, 76–77.

51. From the monarchy through the revolutionary period, the canonical text in English is Batatu 1978. Al-Qarawee 2012, Dodge 2003, and Marr 2004 all provide historical overviews of identity politics in Iraq over the last century.

52. Batatu 1978; Marr 1970.

53. Al-Qarawee 2012; Batatu 1978; Marr 1985.

54. Batatu 1978.

55. Sky 2011.

control meant there was a limit to how much the United States could do to accelerate institution building or keep Maliki from circumventing institutions that needed his deference to gain credibility.⁵⁶

Nationalism also determined the severity of the sovereignty dilemma, which ultimately limited how long the United States could offer state-building assistance in Iraq. As with the sectarian and institutional dilemmas, the United States recognized the underlying tension of the sovereignty dilemma from the beginning of the war. But as long as the sectarian dilemma dominated in 2006, the sovereignty dilemma was latent. It was when the violence plummeted and nationalism re-emerged that the exigencies of Iraqi politics forced Sunni and Shia parties to push the United States for an explicit withdrawal timeline.⁵⁷ The negotiations centered on issues of sovereignty, with Maliki explaining in the initial phases of the negotiation in June 2008, “We have reached a dead end, because when we started the talks, we found that the US demands hugely infringe on the sovereignty of Iraq, and this we can never accept.”⁵⁸

Eventually the agreement did have a timeline for US withdrawal, forcing Iraq’s leaders to consider what incentives they would face after withdrawal. Following the close 2010 election, Maliki subverted the state to punish political threats, particularly among Sunnis.⁵⁹ ISIS reemerged and sectarian violence spiked once again. In terms of our theory, the period of nationalism was too short for state building to occur at the level required to constrain Maliki. The prospect of revitalizing sectarian patronage networks was hard to avoid in the absence of a heavy US footprint.

More specifically, in terms of the model’s parameters, the depth of Iraq’s sectarian hostilities in 2006 (β), alongside the availability of a meaningful if historically contested national Iraqi identity (α), made the nation-building path particularly salient for leader incentives, and success particularly sensitive to the sovereignty dilemma. In such conditions, committing more forces (up to \hat{f}_c) to transform a conflict in the face of a sectarian dilemma will increase the prospects for sustainable peace, as in Propositions 3 and 4. As a result, scenario 2 is more likely to succeed than 1, and 4 more likely than 3. Similarly, the more severe the exit dilemmas, the lower the prospects of success, such that moving from left to right in Table 1 increases the prospects of success, as in Proposition 6.⁶⁰ In Iraq’s case, the successful surge was

56. Maliki responded to the weak institutional environment by developing a chain of command personally loyal to him through the Office of the Commander in Chief. Dodge 2012; Filkins 2014.

57. Gordon and Trainor 2012, 528. News reports at the time of the status of forces agreement negotiations reveal that domestic politics put enormous pressure on Iraq’s politicians to get a timeline for withdrawal in the name of Iraq’s sovereignty. See Steven Lee Myers, “Bush, in a Shift, Accepts Concept of Iraq Timeline,” *New York Times*, 19 July 2008; Alissa Rubin and Campbell Robertson, “Iraq Backs Deal that Sets End of US Role,” *New York Times*, 27 September 2008; Marc Lynch, “Bush’s Finest Moment on Iraq: SOFA, not the Surge,” *Foreign Policy*, 18 January 2009.

58. “Talks on US-Iraq Pact at ‘Dead-End,’” *Al-Jazeera*, 13 June 2008.

59. Emma Sky and Harith Al-Qarawee, “Iraqi Sunnistan? Why Separatism Could Rip the Country Apart—Again,” *Foreign Affairs*, 23 January 2013.

60. In the ideal case modeled earlier, the intervener is trying to forge a democracy. However, majoritarianism, which was the goal for Shia parties in Iraq, need not be perceived as democratic in ethnically

consistent with being in the lower row of Table 1. But the possibility that initial success would mask a shift back to sectarianism, as in Proposition 5, and the sensitivity to the institutional and sovereignty dilemmas, meant that the US–Iraq case was in scenario 2. Inclusive institutions did not emerge (i.e., γ was too high for κ to exceed $\hat{\kappa}$) and so could not contain Maliki’s sectarian turn when he saw it as politically beneficial.

The theory applies more broadly, including to situations where social identity is less sensitive to fighting, or leader choices are less sensitive to social identity, than in Iraq. Specifically, under conditions where α , β , and κ^1 are low and the sectarian dilemma not too severe, and the institutional dilemma can be mitigated by a light footprint ($f_c < \hat{f}_c$), high χ can incentivize leader choices of $c = 1$, while α has no impact on ω because the sovereignty dilemma is not triggered. As a result, a light but long intervention, rather than a nation-building strategy, can effectively state build. That the institutional dilemma can depend on the level of intrusion, while the sovereignty dilemma applies only when intrusion is high and intervention has induced nationalism, creates the possibility that the two dimensions interact, and an intervention may face a choice between scenario 2 and scenario 3 in Table 1, with success more likely in the latter.⁶¹

Summarizing these conclusions, for the transformative nation-building channel to successfully enable major state building, the perceived threat to sovereignty must be low (scenario 4). Alternatively, successful state building under a light footprint may be possible without a population that identifies nationally, as this avoids the sovereignty dilemma, and will be particularly attractive if the lighter footprint mitigates the institutional dilemma as well (as in scenario 3).

Four complications arise. First, no intervener is unconstrained in resources, despite the usefulness of that assumption for distilling the factors at play. This is particularly true due to political forces that will question foreign interventions where “core national interests” are not at stake. Indeed, our theory suggests one reason that interventions in scenario 1 occur despite their poor prospects for success: the much costlier option of scenario 2 is also a low-success option.

Second, both dimensions of Table 1 are hard to evaluate *ex ante* and even during the conflict, as the recurring debates in Iraq illustrated. As a result, policy advocates may either seek to minimize the exit dilemmas by limiting the footprint where strong social identities or humanitarian concerns instead call for transforming the conflict; or, conversely, a heavy footprint to settle violence may trigger severe exit dilemmas when a light footprint could avoid them and succeed, or nonintervention could do the least harm.

divided states. In such cases, the sectarian dilemma will be more severe and will require deeper investments and lengthier intervention.

61. Indeed, scenario 3 may even be a better option than scenario 4 where ethnic hostilities are low and the institutional dilemma is particularly sensitive to troop levels, as our discussion of El Salvador soon illustrates; in Table 1 we highlight the case where social identities are important to the conflict (high α , β), centering the importance of nation building and giving scenario 4 the highest prospects for success.

Third, such miscalculations become more likely because exit dilemmas are second-order consequences of success in the face of the sectarian dilemma and they will tend to be discounted, especially in the face of urgent humanitarian or genocidal threats (which may justify intervention but result in scenario 1).

Fourth, [Table 1](#) compares the possibilities for a given depth of ethnic hostilities (α , β , Y), but their variation further complicates easy categorization of conflicts. If hostilities are deep, then in all cases \hat{f}_c will be higher, and prospects for success will be lower (i.e., γ more likely to be below the threshold necessary as in the propositions of our theory). This “third dimension” clearly weighs on whether an intervener will be able to commit the resources necessary to bring violence down or build institutions sufficiently for sustainable peace. As already discussed, this may also affect whether the nation-building channel for peace is more likely to succeed than light-footprint state building.

Our theory’s contribution is therefore not to deliver a one-size-fits-all answer to intervention, but rather to show how the interactions between intervener and local politics map to policy debates and how these complexities generally weigh toward caution. To show how our theory covers intervention types comprehensively, in the following two sections we categorize them first along the dimension of light- and heavy-footprint intervention (i.e., according to the rows of [Table 1](#)), as this is the more prominent and intuitive in accepted typologies. Within those categories, following the described logic, we consider how external hostility (i.e., toward the intervener, as defined by the columns of [Table 1](#)) and internal ethnic hostilities among the groups shape the possibilities for success.

Light-Footprint Interventions

Humanitarian interventions are typically understood as the provision of military and/or material assistance to protect civilians working in existing state institutions or civil society organizations, and they are therefore classified in the light-footprint category. They may succeed in their limited aims if the perceived threat to sovereignty and internal ethnic hostilities are low (scenario 3). However, such operations are not designed to solve the sectarian dilemma and cannot build peace where ethnic divisions are significant because they do not address the root causes of conflict. They will therefore likely fail to prevent conflict recurrence when intervention ends, and this realization often leads to “mission creep,” whereby foreign-imposed regime change (FIRC) becomes the goal of humanitarian intervention.⁶² Our theory suggests that, in cases where violence and ethnic identification are at high levels, FIRCs will fail if they combine maximalist goals—such as replacing a tyrannical regime with an inclusive one—with an insufficient degree of engagement.

62. Kuperman 2013.

Libya is a case in point: the NATO intervention actually increased levels of ethnic violence during and after the overthrow of the Gaddafi regime,⁶³ leading the country to civil war. Libya's institutions were weak following decades of authoritarian rule, and its lootable resources, combined with salient tribal and sectarian divisions, made it a candidate for civil war when the regime was toppled. Violence made ethnic identities more salient, and the strong factionalism created a sectarian dilemma. Given the lack of appetite for an extensive intervention by the international community and the high threat that partial intervention would pose to local leaders, a more prudent strategy in terms of long-term peace might have been not to intervene in the first place. However, the shift in public opinion against the Gaddafi regime and the fear that non-intervention would create a humanitarian crisis provided support for an intervention where the goals were mismatched to the instruments used. Consistent with our theory, leaders responded by playing ethnic politics, pushing the country toward the high violence / ethnic identification equilibrium. This case illustrates the importance of not allowing FIRCs to break down into anarchy and ethnicized violence in already weakly institutionalized environments.

Libya is a paradigmatic case of a FIRC that failed because intervention did not provide enough resources to resolve the sectarian dilemma. Interventions that aim to remove the underlying causes of genocide, civilian repression, and ethnic cleansing must be more extensive than humanitarian-assistance missions. Research has shown that FIRCs are likely to fail if they are limited to decapitating the existing regime with no follow-through.⁶⁴ Our theory is consistent with that evidence. Downes and Monten distinguish between "leadership" and "institutional" FIRCs and find evidence that leadership FIRCs—which are "high-threat, low-intrusion" operations in our typology—do not have positive outcomes.⁶⁵ Leadership FIRCs are often interventions designed to remove a hostile leader, rather than to create democracy. Institutional FIRCs have a better track record, as they represent a bigger investment and higher degree of coercion (heavier footprint) by the intervener, whose task it becomes to build state institutions in the target country to address the sectarian dilemma. Our theory complements previous scholarship by pointing out that the success or failure of institutional FIRCs may hinge on how the institutional and sovereignty dilemmas are addressed by the intervener.

The conflict in Afghanistan offers a useful example of an institutional FIRC. Following the initial defeat of the Taliban, the US intervention aimed to install and support a government with relatively high inclusiveness to accommodate Afghanistan's fractious social fabric.⁶⁶ That intervention has failed to achieve its goals, at least partly because of a particularly severe institutional dilemma matched

63. Kuperman 2013.

64. Downes and Monten 2013.

65. *Ibid.*

66. On the fragmentation of Afghanistan, see Rubin 2002. On nation building in Afghanistan, see Dobbins 2008.

to a particularly low institutional starting point.⁶⁷ While a coalition of forces participated in this intervention, the local population identifies the International Security Assistance Force with American interests, and local leaders collaborating with the coalition are perceived as puppets. At the same time, while Afghanistan's α is low, there is resistance to occupation, not due to strong nationalism but rather because no faction can represent the national identity, and therefore intervention is unable to shift the country to the national equilibrium. Instead, intervention is perceived as intensifying ethnic/tribal antagonisms over territory or resources, and leaders use the resources provided by the intervener to increase the material benefits and status of their own group. This case reveals that the institutional dilemma can emerge even if there is no nationalist transformation and the sovereignty dilemma does not arise. It also demonstrates how "light footprint" should be defined relative to the parameters of the conflict: even when US forces briefly approached 100,000 in 2010 and 2011 (they were generally below 40,000 for most of the conflict), they could not reach the level required to transform the conflict (\hat{f}_c), perpetuating the scenario 1 dynamics.

Unsurprisingly, success is more likely where ethnic hostilities are not deep, even if nationalism is weak, because this allows the deployment of consent-based multilateral peacekeeping operations. Such operations, usually fielded by the United Nations, are characterized by low fighting capacity; they aim to provide technical assistance, monitoring, and policing of a settlement that is broadly acceptable to most of the local actors. The multilateral nature of the intervention can make it seem more impartial and less threatening to local leaders, and the transformation of resources into public goods is faster, due to the assurances they provide that all local actors are cooperating with the terms of a settlement. It follows that the institutional dilemma will generally be less severe for such interventions, since local leaders are not crowded out, while the lighter footprint avoids the sovereignty dilemma altogether (characterized by the relatively high prospects of success in scenario 3).

Examples of such missions include the UN interventions in El Salvador and Cambodia in the 1990s. Early multidimensional peace operations in El Salvador (ONUSAL) and Cambodia (UNTAC) were decisive in inducing the parties to negotiate over the contours of a postwar settlement, provided assurance that the parties were not violating the terms of the agreement, helped build institutions and run elections, and provided tactical enforcement in response to efforts by spoilers to undermine peace treaties.⁶⁸ The UN's perceived impartiality and the principle of consent were central to the success of these operations.

67. Some of the insights of our model are reflected in an editorial by Max Fisher and Amanda Taub, "Why Afghanistan's War Defies Solutions," *New York Times*, 24 August 2017; and in Craig Whitlock, "Nation-Building in Afghanistan," *Washington Post*, 9 December 2019.

68. Doyle, Johnstone, and Orr 1997; Doyle and Sambanis 2006.

Heavy-Footprint Interventions

Whereas light-footprint interventions are unlikely to induce a nation-building path to peace and therefore risk the dangers of power vacuums even when limited humanitarian and FIRC goals are achieved, heavy-footprint interventions face the opposite risk: rejection by the very state they seek to strengthen. Where ethnic hostilities are high, heavy-footprint interventions may take the form of neo-trusteeship, particularly when elites think they could win the contest via continued fighting, or when institutions are so weak that even cooperative elites cannot deliver public goods. The risk of “public bads” associated with such failed states generated a swell of interventions after the end of the Cold War.⁶⁹ Many failed because they were not actually resourced as heavy-footprint interventions, but some succeeded. In contrast to the Iraq case already discussed, where the sovereignty threat perception was high (scenario 2), these interventions were generally able to minimize the contradictions of external assistance by carefully calibrating their mandate and by managing the sovereignty threat created by the intervener (scenario 4).

The identity of the intervener is a crucial determinant of the degree to which neo-trusteeship will be perceived as a threat to local actors. Multilateral interventions led by the UN pose less risk than unilateral trusteehips. The UN’s involvement in Eastern Slavonia (Croatia) and in the municipality of Brčko (Bosnia) are successful examples of neo-trusteeship in which a heavy footprint did not trigger the sovereignty dilemma.⁷⁰ As in all cases, the model cannot explain every factor that contributed to this outcome. The NATO-imposed outcome of Yugoslavia’s wars, international recognition of Bosnia and Croatia, defeat of Serb radicals, and an ethnic power-sharing pact codified in the Dayton Accords all contributed to ending ethnic war.

Multilateral peacekeeping is usually considered a light-footprint intervention in terms of absolute troop levels, but can still work even if ethnic hostilities are high and ethnic identities are salient. Under these conditions, our theory suggests that peacekeepers must work to reduce ethnic fighting, to allow national identities to take hold (i.e., to reach our threshold of heavy footprint, \hat{f}_c , which may depend on strategy and capabilities as well as troop levels). This has important implications for ongoing debates about the proper use of force in consent-based peacekeeping missions. Mission effectiveness depends on the ability to control spoilers and extremists who will try to use violence to undermine the peace.⁷¹ If such violence is not contained, the population will not have the opportunity to shift to a national-identity equilibrium. The key question is whether a peacekeeping force can use violence to target extremists without increasing the perceived threat to local actors who are committed to peace. Toward this end UN interventions are more effective than unilateral efforts,

69. Fearon and Laitin 2004.

70. Doyle and Sambanis 2006.

71. Ibid.

and recent research suggests that they contribute to state-building imperatives without supplanting the state.⁷²

Examples of such missions include the UN interventions in Côte d'Ivoire (2004 to 2017) and Liberia (2003 to 2018). The mandate of the UN Mission in Liberia, authorized under Chapter VII of the UN Charter, included the restoration of law and order after civil war had divided the country ethnically. Blair shows that the UN was effective in inducing citizens to rely on the rule of law, shoring up support for state actors without crowding them out.⁷³ By virtue of its greater impartiality and legitimacy, UN peace building may succeed, as long as resources are adequately matched to the underlying peace-building ecology.

While UN missions with a heavy footprint are more likely to be in scenario 4 rather than 2, compared to other interveners, they are more likely to face coordination problems and other constraints due to a slow-moving bureaucracy and the political constraints of the UN Security Council. The intervention diagnostic is therefore particularly relevant for the United States, which is most able to achieve the footprint necessary for long-term peace through its resources and military capabilities and yet clearly runs a high risk of inducing high sovereignty threat in ethnic conflicts, especially in the Middle East.

The civil war in Syria is therefore another critical application of our theory. Syria's civil war bears a number of similarities to Iraq's in 2006 and 2014. Indeed, not only is it roughly the same ethno-sectarian groups, it can be considered a single battlefield, drawing in the same regional interests. A very high f_c and some form of neo-trusteeship would have been necessary to solve an intense sectarian dilemma in Syria: the high degree of factionalism, exceedingly high levels of hostility (as measured by the total number of deaths and displacements), the sectarian nature of the fighting, and devastated infrastructure and economy all suggested that the fighting could end only with a large boots-on-the-ground intervention and occupation.

By all accounts, Syrian leaders catered only to their narrow ethnic or sectarian constituencies (low μ), and high levels of ethnic violence had weakened national identities. The debate on whether the United States should have intervened reflects the centrality of the sectarian dilemma. President Obama likely recognized that an Iraq-style mission was needed to have a lasting impact (in our terms, overcome the sectarian dilemma). His ambivalence was informed by the Iraq experience, where intervention failed because there was no feasible exit strategy that could guarantee sufficient progress in state building before the sovereignty dilemma kicked in; and because the institutional dilemma limited American initiatives to support local leaders. Obama's position reflected the inevitability that heavy intervention and occupation of Syria, even if it could reduce violence, would generate "high threat" for any otherwise promising Syrian leaders and complicate any mission for durable peace. By contrast, advocates of intervention focused on the need to reduce civilian suffering,

72. Blair 2015.

73. Blair 2019.

raising the prospect of taking on the sectarian dilemma. What complicates this case further—and is a parameter outside of the scope of our model—was the specter of a proxy war with Russia over competing interests in Syria and in the broader regional war. Research has shown that competing interventions lengthen the duration and severity of civil wars; with Russia and Iran opposing the United States, the prospects of a successful nation-building intervention that could create self-enforcing peace in Syria were smaller than in Iraq.

The Syria example highlights a key implication of our theory, which is that non-intervention is the more prudent strategy unless intervention can be done in such a way as to defuse threats to sovereignty and encourage collaboration by local leaders, shifting the conflict from scenario 1 or 2 to scenario 4 in [Table 1](#). While President Obama's caution about a potential scenario 2 outcome was justified under our theory, the humanitarian cost requires exploring the possibilities for multi-lateral solutions compatible with scenario 4. Alternatively, accepting the impossibility of nation building, a powerful internever can use force to try to lock in a balance of power that does not resolve the sectarian dilemma but stops the violence, as was the case in the Dayton Accords that ended Bosnia's war. Regional autonomy or loose confederal arrangements could be considered in such cases, though these solutions also have their own sets of limitations and make it hard for inclusive institutions to emerge.

Conclusion

Faced with the challenge of intervening to end atrocities of ethno-sectarian war, policy-makers should consider the implications of how social identities, leader incentives, and institutional development interact. When leaders have come to power by fighting for their ethnic groups, it takes a heavy intervention to change their incentives, and such a change is more likely if the population shifts from ethnic to national identification. Inclusive leadership becomes possible when national identities grow stronger. A key question is whether intervention can achieve this shift from ethnic to national identification or whether it is likely to distort local politics, marginalize local leaders, and induce nationalist opposition to foreign presence. Successful peace building does not only depend on resource provision, technical know-how, enforcement capacity, or power-brokering prowess; and even well-resourced interventions that are sensitive to the balancing act of forging incentives for self-enforcing peace can fail, limited by their lack of legitimacy in the central task of building self-governance.

Our theory provides a sobering assessment of the prospects for third-party-led nation building after ethno-sectarian war. Prospects of success turn on the parameters that define the severity of the dilemmas we have explored. Prospects are poor when ethnic identity is easily induced by fighting, when the occupier's efforts to create institutions make local leaders look like puppets, and when coercive occupation evokes colonialism. This does not bode well for US-led democracy promotion through forcible regime change and other versions of third-party-led political

engineering in countries with low *ex ante* levels of institutional development and salient ethno-sectarian divisions. Taking identities seriously helps us see that the intervener's identity can interfere with state building in ways that are unpredictable because they depend on specific features of the afflicted country's history.

Transformative interventions can succeed if they balance committing sufficient resources to contain violence and minimizing perceived threats to leaders who commit to inclusive politics—but their control is limited, and the fate of the intervention rightfully depends on the choices of local leaders. The art of nation building after ethnic war relies on managing the distortions that an external force naturally introduces into local politics and finding a way for the local peace builder to replace the peace enforcer. At a minimum this means that interventions should not be led by a power that stands for historical exploitation—which points toward multilateral efforts. In addition, there must be a cohesive national identity that leaders can appeal to. That identity must be open and inclusive, not tied to a specific ethnic group that marginalizes the others. Under such conditions, nationalism can generate more cross-ethnic support for leaders who prove their competency and commitment to providing public goods that benefit all groups. If leaders who govern by building inclusive institutions cannot appeal to a nationalism that facilitates cross-ethnic support, then the nation-building channel will close, and it cannot support state building.

Given the challenges of a heavy-handed intervention, an alternative strategy is to intervene with a lighter footprint. These interventions avoid triggering the sovereignty dilemma, but they will work only if they can make progress on state building without the help of a unifying national identity. Their goal should be to develop institutions with the help of local elites more incrementally, without an impending withdrawal date, so that leaders have assurances that violence will not recur as they take risks for equality and unity. Such state-building models are likely to be successful only where hostilities are already mild; where the intervener can neutralize spoilers; or where leaders are already predisposed to cooperate and can legitimately claim the credit for success, making institutional commitments credible in the eyes of an otherwise divided population. Where ethnic hostilities remain high and the institutional dilemma is strong, then even a long (yet light-footprint) intervention can be ineffective, as it was in Afghanistan, and a short intervention can make things worse by creating a power vacuum, as it did in Libya. In such cases, or where proxy fighting or cross-border sanctuaries make any nation-building effort tenuous, as in Syria, a best option may be to try to freeze the conflict in place to avoid further bloodshed or mass displacement. This strategy will not solve the sectarian dilemma, but could answer the humanitarian imperative until exogenous complicating factors (regional conflict dynamics) have shifted.

The intervention “diagnostic” we propose integrates parallel debates over commitment of resources to stop ethno-sectarian violence and concern for the distortion of local politics and lack of legitimacy in shaping them. As long as evolving norms of intervention favor early engagement to stop civilian suffering before the parties have fought to a stalemate, the sectarian dilemma will arise. Recent experiences in

Iraq, Afghanistan, Libya, and Syria suggest that the challenge will persist, and different approaches have all yielded costly failures.

By highlighting both how committing sufficient resources can overcome the sectarian dilemma “on the way in” and the limits for successful exit strategies “on the way out,” our theory offers a unified framework for interventionists and skeptics to constructively engage each other on the same terms. Rather than talking past each other by focusing on one dynamic or the other, the two sides should focus on the severity of the interlocking dilemmas and the corresponding conditions for success of any peace-building intervention, alongside practical and normative considerations. As long as the prospect of mass atrocities induces calls to intervene, a better understanding of these second- and third-order consequences should shape the policy of whether to do so, and how.

Supplementary Material

Supplementary material for this article is available at <https://doi.org/10.1017/S0020818321000266>.

References

- Al-Qarawee, Harith. 2012. *Imagining the Nation: Nationalism, Sectarianism and Socio-political Conflict in Iraq*. Rossendale Books
- Batatu, Hannah. 1978. *The Old Social Classes and the Revolutionary Movements of Iraq: A Study of Iraq's Old Landed and Commercial Classes and of Its Communists, Ba'thists and Free Officers*. Princeton University Press.
- Beath, Andrew, Fotini Christia, and Ruben Enikolopov. 2012. Winning Hearts and Minds Through Development? Evidence from a Field Experiment in Afghanistan. Policy Research Working Paper 6129, World Bank.
- Besley, Timothy, and Torsten Persson. 2010. State Capacity, Conflict, and Development. *Econometrica* 78 (1):1–34.
- Biddle, Stephen, Jeffrey A. Friedman, and Jacob Shapiro. 2012. Testing the Surge: Why Did Violence Decline in Iraq in 2007? *International Security* 37 (1):7–40.
- Biddle, Stephen, Michael E. O'Hanlon, and Kenneth M. Pollack. 2008. How to Leave a Stable Iraq: Building on Progress. *Foreign Affairs* 87 (5):40–58.
- Blair, Robert. 2015. Four Essays on Peacebuilding: The Consolidation of State Authority. PhD dissertation, Yale University.
- Blair, Robert. 2019. International Intervention and the Rule of Law After Civil War: Evidence from Liberia. *International Organization* 73 (2):365–98.
- Brownlee, Jason. 2007. Can America Nation-Build? *World Politics* 59 (2):314–40.
- Call, Charles T. 2008. Ending Wars, Building States. In *Building States to Build Peace*, edited by Charles T. Call and Vanessa Wyeth, 365–88. Lynne Rienner.
- Chesterman, Simon. 2004. *You, the People: The United Nations, Transitional Administration, and Statebuilding*. Oxford University Press.
- Dietrich, Simone, and Matthew S. Winters. 2015. Foreign Aid and Government Legitimacy. *Journal of Experimental Political Science* 2 (2):164–71.

- Dobbins, James F. 2008. *After the Taliban: Nation-Building in Afghanistan*. Potomac Books.
- Dobbins, James, Seth Jones, Keith Crane, and Beth Cole DeGrasse. 2007. *The Beginner's Guide to Nation-Building*. RAND.
- Dobbins, James, John G. McGinn, Keith Crane, Seth G. Jones, Rollie Lal, Andrew Rathmell, Rachel M. Swanger, and Anga R. Timilsina. 2003. *America's Role in Nation-Building: From Germany to Iraq*. RAND.
- Dodge, Toby. 2003. *Inventing Iraq: The Failure of Nation Building and a History Denied*. Columbia University Press.
- Dodge, Toby. 2012. Iraq's Road Back to Dictatorship. *Survival: Global Politics and Strategy* 54 (3):147–68.
- Downes, Alexander B., and Jonathan Monten. 2013. Forced to Be Free? Why Foreign-Imposed Regime Change Rarely Leads to Democratization. *International Security* 37 (4):90–131.
- Doyle, Michael, Ian Johnstone, and Robert C. Orr. 1997. *Keeping the Peace: Lessons from Multidimensional UN Operations in Cambodia and El Salvador*. Cambridge University Press.
- Doyle, Michael, and Nicholas Sambanis. 2006. *Making War and Building Peace: United Nations Peace Operations*. Princeton University Press.
- Edelstein, David M. 2009. Foreign Militaries, Sustainable Institutions, and Postwar Statebuilding. In *The Dilemmas of Statebuilding: Confronting the Contradictions of Postwar Peace Operations*, edited by Roland Paris and Timothy Sisk, 81–103. Routledge.
- Fearon, James, and David Laitin. 2004. Neotrusteeship and the Problem of Weak States. *International Security* 28 (4):5–43.
- Filkins, Dexter. 2014. What We Left Behind. *New Yorker*, 28 April.
- Fukuyama, Francis. 2004. Nation-Building 101. *The Atlantic*, January/February.
- Fukuyama, Francis. 2005. "Stateness" First. *Journal of Democracy* 16 (1):84–88.
- Gordon, Michael R., and Bernard E. Trainor. 2012. *The Endgame: The Inside Story of the Struggle for Iraq, from George W. Bush to Barack Obama*. Pantheon Books.
- Hechter, Michael. 2001. *Containing Nationalism*. Oxford University Press.
- Hechter, Michael. 2013. *Alien Rule*. Cambridge University Press.
- Kaufmann, Chaim. 1996. Possible and Impossible Solutions to Ethnic Civil Wars. *International Security* 20 (4):136–75.
- Kuperman, Alan. 2013. A Model Humanitarian Intervention? Reassessing NATO's Libya Campaign. *International Security* 38 (1):105–36.
- Lake, David. 2016. *The Statebuilder's Dilemma: On the Limits of Foreign Intervention*. Cornell University Press.
- Lawrence, Adria. 2013. *Imperial Rule and the Politics of Nationalism: Anti-Colonial Protest in the French Empire*. Cambridge University Press.
- Marr, Phebe. 1970. Iraq's Leadership Dilemma: A Study in Leadership Trends, 1948–1968. *Middle East Journal* 24 (3):283–301.
- Marr, Phebe. 1985. The Development of a Nationalist Ideology in Iraq, 1920–1941. *Muslim World* 75 (2): 85–101.
- Marr, Phebe. 2004. *A Modern History of Iraq*. Westview Press.
- Matanock, Aila, and Adam Lichtenheld. 2015. How Does International Intervention Work? Mechanisms for Securing Peace Settlements in Civil Conflicts. Unpublished manuscript, University of California, Berkeley.
- Myerson, Roger B. 2009. A Field Manual for the Cradle of Civilization: Theory of Leadership and Lessons of Iraq. *Journal of Conflict Resolution* 53 (6):470–482.
- Mylonas, Harris. 2012. *The Politics of Nation-Building: Making Co-nationals, Refugees, and Minorities*. Cambridge University Press.
- Paris, Roland. 2004. *At War's End: Building Peace After Civil Conflict*. Cambridge University Press.
- Paris, Roland. 2014. The "Responsibility to Protect" and the Structural Problems of Preventive Humanitarian Intervention. *International Peacekeeping* 21 (5):569–603.
- Ricks, Tom. 2009. *The Gamble: General David Petraeus and the American Military Adventure in Iraq, 2006–2008*. Penguin Press.

- Rubin, Barnett. 2002. *The Fragmentation of Afghanistan: State Formation and Collapse in the International System*. Yale University Press.
- Rubin, Barnett. 2008. The Politics of Security in Postconflict Statebuilding. In *Building States to Build Peace*, edited by Charles T. Call and Vanessa Wyeth, 25–47. Lynne Rienner.
- Sambanis, Nicholas, and Moses Shayo. 2013. Social Identification and Ethnic Conflict. *American Political Science Review* 107 (2):294–325.
- Sambanis, Nicholas, Stergios Skaperdas, and William Wolforth. 2015. Nation-Building Through War. *American Political Science Review* 109 (2):279–96.
- Sisk, Timothy. 2013. *Statebuilding: Consolidating Peace After Civil War*. Polity Press.
- Sky, Emma. 2011. From Surge to Sovereignty. *Foreign Affairs* 90 (2):2–12.
- Smith, Jean Edward. 2016. *Bush*. Simon and Schuster.
- Visser, Reidar. 2011. *A Responsible End? The United States and the Iraqi Transition, 2005–2010*. Just World Books.
- Walter, Barbara F. 1997. The Critical Barrier to Civil War Settlement. *International Organization* 51 (3): 335–64.
- Wimmer, Andreas. 2018. *Nation Building: Why Some Countries Come Together While Others Fall Apart*. Princeton University Press.
- Woodward, Bob. 2008. *The War Within: A Secret White House History 2006–2008*. Simon and Schuster.

Authors

Kevin Russell is an independent scholar and can be reached at kevin.d.russell@gmail.com.

Nicholas Sambanis is Presidential Distinguished Professor of Political Science and Director of the Identity and Conflict Lab in the Department of Political Science, University of Pennsylvania, Philadelphia. He can be reached at sambanis@upenn.edu.

Acknowledgments

For useful comments, we thank Kate Baldwin, Rob Blair, Lars-Erik Cederman, Alex Debs, Adria Lawrence, Brendan O’Leary, Jonah Schulhofer-Wohl, Stergios Skaperdas, Andreas Wimmer, and the editors of this journal. We especially thank Seok-ju Cho, Tasos Kalandrakis, Jason Davis, and Zachary Taylor for help with the formalization of the argument. This paper was presented at the University of Pennsylvania, Yale University, Uppsala University, and the Hebrew University of Jerusalem.

Key Words

Intervention; sectarianism; nation building; peace building

Date Received: January 9, 2020; Date accepted: January 5, 2021