International Studies Quarterly (2021) **0**, 1–14

Subgroup Differences in Implicit Associations and Explicit Attitudes during Wartime

Aaron Erlich
McGill University, Canada

and

Calvin Garner
University of Washington, USA

When their country is at war, individuals express support for their government and hostility toward the foreign adversary, leading to the "rally 'round the flag" effect. What is less understood is how, during a rally, ethnic identity and proximity to conflict relate to attitudes toward the home state and the adversary. Moreover, individuals may feel pressure to answer patriotically when asked about the conflict, particularly individuals who share an ethnic identity with the majority population of the foreign adversary, leading to biased measures of opinion. We study these dynamics in the context of Ukraine's ongoing war with Russia, comparing responses from self-identified ethnic Ukrainians and Russians in four cities in Ukraine. Using a lab-based implicit association test (IAT) and survey with 600 respondents, we examine whether respondents' implicit biases, reflexive preferences that are hard to manipulate, match their explicitly stated preferences for either Ukraine or Russia. We find that, on average, ethnic Ukrainians and Russians in Ukraine are explicitly and implicitly pro-Ukraine, although we observe slightly lower levels of pro-Ukraine bias among ethnic Russians. We also find that 70 percent of those who are implicitly pro-Russia are explicitly neutral or pro-Ukraine, highlighting the need to study implicit associations in sensitive settings.

Cuando un país está en guerra, la población expresa su apoyo al gobierno y hostilidad hacia el adversario extranjero, lo cual produce el efecto de "agruparse alrededor de la la bandera" ["rally 'round the flag"]. Lo que no se comprende bien es la manera en que, durante una agrupación alrededor de la bandera, la identidad étnica y la proximidad al conflicto se relacionan con ciertas actitudes hacia el estado local y el adversario. Además, las personas pueden sentirse presionadas para responder de forma patriótica cuando se les pregunta acerca del conflicto, en particular aquellas que comparten una identidad étnica con la mayoría de la población del adversario extranjero, algo que da lugar a medidas de opinión sesgadas. Estudiamos estas dinámicas en el contexto de la guerra actual entre Ucrania y Rusia, comparando respuestas de ucranianos y rusos autoidentificados étnicamente en cuatro ciudades de Ucrania. Utilizando una prueba de asociación implícita (implicit association test, IAT) de laboratorio y una encuesta con 600 participantes, examinamos si los sesgos implícitos (preferencias reflexivas que son difíciles de manipular) de los encuestados coinciden con sus preferencias expresadas, ya sea a favor de Ucrania o de Rusia. Descubrimos que, en promedio, las personas de origen étnico ucraniano y ruso en Ucrania están explícita e implícitamente a favor de Ucrania, aunque observamos niveles levemente más bajos de sesgo a favor de Ucrania en personas de origen étnico ruso. Además, observamos que el 70 percent de aquellas personas que están implícitamente a favor de Rusia están explícitamente a favor de Ucrania o tienen una postura neutral, lo cual resalta la necesidad de estudiar asociaciones implícitas en entornos sensibles.

Lorsque leur pays est en guerre, les individus expriment leur soutien pour leur gouvernement et leur hostilité envers l'opposant étranger, ce qui mène à un effet de « ralliement autour du drapeau ». Cependant, ce qui est moins compris, c'est la mesure dans laquelle l'identité ethnique et la proximité du conflit ont un lien avec les attitudes envers l'État de résidence et l'opposant durant ce ralliement. De plus, les individus peuvent ressentir une pression les poussant à répondre patriotiquement lorsqu'ils sont interrogés sur le conflit, particulièrement les individus qui partagent l'identité ethnique de la majorité de la population de l'opposant étranger, ce qui mène à des mesures d'opinion biaisées. Nous étudions ces dynamiques dans le contexte du conflit continu entre l'Ukraine et la Russie en comparant les réponses d'individus qui s'identifient d'eux-mêmes ethniquement en tant que Russes ou qu'Ukrainiens dans quatre villes d'Ukraine. Nous utilisons un test d'association implicite mené en laboratoire et une enquête sur 600 participants pour examiner si les préjugés implicites des participants, des préférences réflexives difficiles à manipuler, correspondent à leurs préférences explicitement déclarées envers l'Ukraine ou la Russie. Nous avons constaté qu'en moyenne, les Ukrainiens et Russes ethniques vivant en Ukraine étaient explicitement et implicitement pro-Ukraine, bien que nous ayons observé des niveaux de préférences pro-Ukraine légèrement inférieurs chez les Russes ethniques. Nous avons également découvert que 70 percent des personnes qui étaient implicitement pro-Russie étaient explicitement neutres ou pro-Ukraine, ce qui met en évidence le besoin d'étudier les associations implicites dans les environnements sensibles.

Aaron Erlich is an Assistant Professor of Political Science at McGill University. His research interests include the impact of political information, measurement, democratization, and experimental design. Previous work has appeared in *American Political Science Review*, *Journal of Politics,* and *Comparative Political Studies,* and other journals.

Calvin Garner is a PhD candidate in the Department of Political Science at the University of Washington. He has substantive interests in the study of behavior, institutions, and conflict in authoritarian, democratizing, and newly democratic regimes. He has a regional focus on countries of the former Soviet Union.

Author's note: We thank the editors at *International Studies Quarterly*, two anonymous reviewers, Tymofii Brik, Megan Bradley, Michael Burch, Sarah Bush, Henry Hale, Egor Lazarev, Volodomyr Kulyk, Matthew Light, James Long, Krzysztof Pelc, Lee Seymour, Gavin Slade, Geoff Wallace, Nora Webb Williams, and seminar participants at the Centre for International Peace and Security Studies (CIPPS) at McGill University; the Severyns Ravenholt Seminar in Comparative Politics at the University of Washington; the Centre for European, Russian, and Eurasian Studies (CERES) at the University of Toronto; the Midwest Political Science Association (MPSA) 2018 Annual Meetings; Urban Space (Kyiv); and Temple University. All errors remain ours. The authors collected these data in collaboration with the National Democratic Institute, which funded the research. The analysis and conclusions contained here solely represent those of the authors. The research comes under either University of Washington IRB protocol #00007684 or McGill University REB protocol #390-0217. The data underlying this article are available on the ISQ Dataverse at https://dataverse.harvard.edu/dataverse/isq.

Erlich, Aaron, and Calvin Garner. (2021) Subgroup Differences in Implicit Associations and Explicit Attitudes during Wartime. *International Studies Quarterly,* doi: 10.1093/isq/sqab005

A well-developed international relations literature demonstrates that when a country is at war, its population will "rally ′round the flag." Prior work on diversionary war and the "rally effect" argues that the new, highly salient threat posed by a wartime adversary causes a surge in individuals' support for their home state and its leaders (Mueller 1973; Russett 1990; Davies 2002; Lai and Reiter 2005), anger, and support for military response to the threat (Lambert et al. 2010). In this article, we seek to improve our understanding of the rally effect by engaging with three questions about attitudes in the context of a wartime rally.

First, how do ethnic ties to the aggressor state relate to attitudes toward the home state and the aggressor? On the one hand, a comparative politics literature on ethnicity and conflict suggests that the outbreak of a conflict can create or exacerbate ethnic divisions within a given state and prolong fighting (Kaufmann 1996; Fearon 2004; Buhaug, Cederman, and Gleditsch 2014). If ethnic identity becomes a salient division in a society at war, members of an ethnic minority group may express attitudes about that war that differ from those of the majority ethnic group, particularly when the minority group has ties to the wartime adversary (Mylonas 2012). Suspicions that minority populations are disloyal can exacerbate tensions between these populations and their governments and, in some cases, lead governments to commit civil and human rights abuses (Wallace 2015). On the other hand, there is evidence that the threat of external conflict strengthens ties to the home state across ethnic groups and forges new, pro-home-state identity. Increased pro-home-state identity may occur because the state becomes more intimately involved in the lives of its citizens to meet the exigencies of war (Herbst 1990; Tilly 1990). A newly salient, hostile out-group raising the importance of civic identity, regardless of ethnicity, may also contribute to higher levels of pro-home-state identity (Saideman and Ayres 2008).

Second, does physical proximity to conflict predict different levels of support for the home state and antipathy toward the aggressor within the context of a rally? There is considerable variation in how individuals experience war on the local level, with these experiences often diverging from higher-level narratives about the conflict (Kalyvas 2006; Wood 2008; Balcells 2012). Spatial variation in the risks associated with the conflict could lead to differences in the salience of the conflict, impacting attitudes toward the home state and the aggressor (Bakke, O'Loughlin, and Ward 2009; Tellez 2019).

Third, can we trust the data? Examining attitudes about the home state and the aggressor during wartime presents a major methodological difficulty because these are the opinions that respondents would most likely feel pressure to falsify. Survey respondents may feel considerable pressure to voice support for their home country and hostility toward the adversary in times of conflict, especially when asked the question by a survey enumerator whom they have just met. This problem is acute when trying to elicit the views of ethnic minority populations who have ties to the wartime adversary. History shows that such subgroups are vulnerable to persecution based on their identity, so they may be ready to overstate their loyalty to the state and their hostility toward the aggressor.[1]

To answer these questions, we examine Ukrainian citizens' attitudes toward Ukraine and Russia in the context of the ongoing conflict between the two countries. The case is an ideal one to study our questions because the conflict between Ukraine and Russia led to a documented "rally effect" in Ukrainian public opinion that saw greater support for Ukraine and hostility toward Russia (Kulyk 2016; Korovkin and Makarin 2019). Ukraine's substantial self-identified ethnic Russian minority could plausibly feel allegiance to Russia and may not feel the same pro-Ukraine sentiment as other respondents. Moreover, the risks associated with the conflict have always been greatest in the east of Ukraine, meaning that spatial variation in the salience of the conflict could lead to regional variation in attitudes toward the home state and the aggressor. Finally, the survey work that documents the rally effect in Ukraine does not explicitly engage with the fact that respondents may feel intrinsic or extrinsic pressure to state pro-Ukraine and anti-Russia views, potentially biasing researchers' estimates of public opinion.

We measured 600 respondents' attitudes toward Ukraine and Russia in four Ukrainian cities in April 2015, roughly a year after the start of hostilities. We used a traditional questionnaire to evaluate explicitly held attitudes toward Ukraine and Russia, then took the difference of these attitudes to calculate each respondent's *explicit bias* in favor of one country or the other. Respondents also took a computer-based *implicit association* test (IAT) that revealed their automatic, reflexive bias in favor of one of the two countries. Implicit associations are valuable to researchers because they are often strong predictors of behavior, sometimes stronger than explicitly held views (Roccato and Zogmaister 2010). Implicit associations do not always correlate with explicitly stated attitudes, a phenomenon called *dissociation*. Dissociation is particularly likely in settings in which respondents may feel reluctant to state their opinion frankly or when they may have multiple, crosscutting loyalties to consider in forming an explicit opinion. If individuals are dissociating, then traditional survey research, which records only an explicit attitude, does not tell the whole story. Our research design allows us to test for congruence of implicit and explicit bias, as well as both *pro-Ukraine dissociation*, i.e., holding implicit preferences for Russia but voicing a pro-Ukraine explicit bias, and *pro-Russia dissociation*, in which respondents state preference for Russia when they actually have an implicit bias for Ukraine.

To preview our results, we find that, on average, members of all ethnic identity groups, including ethnic Russians, in all four cities express implicit and explicit preference for Ukraine over Russia, although in some cases not at levels of statistical significance. That said, self-reported ethnicity does serve as a meaningful predictor of relative levels of pro-Ukraine sentiment: respondents who identify as ethnic Russians express a weaker preference for Ukraine than do those who identify as ethnic Ukrainians. We find no consistent support for conflict proximity as a driver of bias. We use nationally representative survey data and weights based on city-level election results to check for robustness.

Comparing the results of the IAT and the explicit questionnaire, we find moderate evidence of attitude dissociation, with just under two in five respondents reporting explicit biases that are at odds with their implicit biases. Ethnicity is less important than the nature of respondents' implicit biases in predicting dissociation; dissociation is only somewhat more common among ethnic Russian respondents than among ethnic Ukrainians, but those who are implicitly pro-Russia are more than twice as likely to dissociate than those who are implicitly pro-Ukraine. Still, to our surprise, we found a large number of respondents dissociating in favor of Russia—that is, having a pro-Ukraine implicit bias

[1] Examples include persecutions of "fifth column" populations, such as the internment of US citizens of Japanese descent during World War II and the ethnic cleansing in the Soviet Union before, during, and after World War II (see Martin 1998).

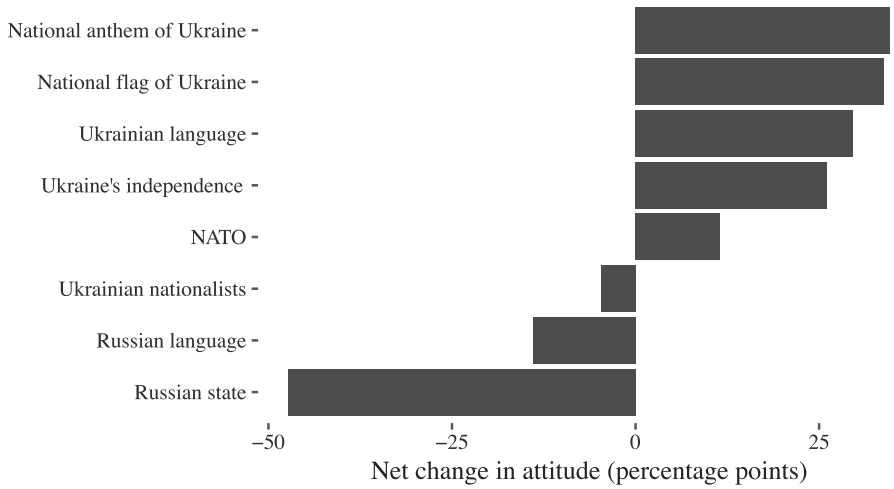

Figure 1. Data from September 2014 reported in Kulyk (2016) show a substantial rally effect as measured by respondents' reported changes in attitudes toward subjects associated with Ukraine and Russia. The values in the figure are the differences between the percentages of respondents who said that their attitude toward a subject improved "a lot" or "a little" and the percentages who said their attitude worsened "a lot" or "a little."

but stating a neutral or pro-Russia explicit bias. This pro-Russia dissociation runs counter to what one would expect from a classical view of attitude dissociation, i.e., that the war would induce those who are implicitly pro-Russia to conceal this implicit association by stating pro-Ukraine explicit attitudes, while those who were implicitly pro-Ukraine would be free to express this opinion. This unexpected finding merits further study in its own right and in other contexts.

This study makes important advances in our knowledge about conflict, identity, and research methods. First, it dispels the notion that ethnic minorities will necessarily feel allegiance to their purported external ethnic homeland in times of conflict. Second, it shows that ethnic identity is an important predictor of the degree of home-state bias in a society where a rally effect has occurred. Third, it documents how those who implicitly favor the wartime adversary dissociate at a higher rate than those who are implicitly biased in favor of the home state. This last finding underscores the value in studying implicit associations as well as explicit attitudes, building on work that has used the IAT to measure potentially sensitive opinions in similarly challenging settings, such as attitudes toward the leader of an authoritarian regime (Truex and Tavana 2019). While others have deployed tools to measure social desirability bias or other forms of deliberate attitude falsification in war zones (see Lyall, Blair, and Imai 2013; Blair, Imai, and Lyall 2014), we believe we are among the first to study the congruence or dissociation of implicit and explicit biases toward the belligerents in a military conflict. Had we only looked at explicit survey responses, we might have misstated the role of ethnicity in predicting attitudes about the belligerents and we would have completely missed the sizable proportion of our sample that dissociated *in favor* of the adversary.

Case Selection

After several months of political turmoil in Ukraine, Russian forces invaded and seized the southeastern Ukrainian region of Crimea in February and March 2014. Shortly thereafter, the Russian military began to provide support to an insurgent movement fighting government forces in eastern Ukraine. In the summer of 2014, Russian troops invaded eastern Ukraine to support these insurgents in their fight against the Ukrainian military. A ceasefire in September 2014 was abandoned amid heavy fighting, but was followed by a second ceasefire signed in February 2015. This second ceasefire has held and led to the establishment of a "line of contact" monitored by the Organization for Security and Cooperation in Europe (OSCE), although violations have occurred frequently, and both civilian deaths and military casualties have continued.

There are at least three reasons why this conflict provides a good case for study. First, the conflict led to a pro-Ukraine, anti-Russia rally in public opinion (see figure 1). In September 2014, Ukrainian survey respondents reported that their attitudes toward the Ukrainian national anthem, flag, language, and Ukraine's independence had improved by at least 25 percentage points from where they were before the conflict onset (Kulyk 2016, 599). Attitudes toward the Russian state dropped by nearly 50 percentage points. Nationally representative public opinion data from before and after the conflict confirm respondents' recollections by showing a sharp deterioration in public attitudes toward Russia after the start of the war.[2]

Second, Ukraine's large ethnic Russian minority is a compelling group to study within the context of a wartime rally. On the one hand, relations between ethnic Ukrainians and ethnic Russians were peaceful (Posen 1993), identity was not politicized to the degree it was in other post-Soviet states, and Russians did not suffer civic discrimination after independence (see Fournier 2002). Further, in the first decade of independence, politics hinged less on ethnic identity cleavages than it did on left–right ideological divides, arguments over economic integration with Europe versus the Russia-led Commonwealth of Independent States (CIS), attitudes toward the Ukrainian state, and elite politics (Shevel 2002; Abdelal 2005; D'Anieri 2007; Hale 2011). On the other hand, the Russian government's messaging and media framing at the start of the war, much of which broke through to Ukrainian audiences, framed the conflict as one where radical, neo-Nazi elements in the Ukrainian state

[2] Korovkin and Makarin (2019) further discuss the way that Ukrainians' attitudes toward Russia deteriorated after the conflict's onset.

Figure 2. A map of Ukraine showing the four cities in which IATs were conducted. Shaded regions indicate Crimea, which was occupied and then annexed by Russia in March 2014, and regions partially controlled by separatist forces at the time of the study.

and society committed atrocities against ethnic Russians (Lankina and Watanabe 2017). This appeal could have fragmented society along ethnic lines, particularly in light of the fact that Ukrainian national identity has been contested and weakly defined in the post-independence period (Korostelina 2013).

Research conducted since the conflict's onset appears to confirm the low political salience of ethnic identity in Ukraine. Frye (2015) shows that whether candidates for office were ethnically Russian or spoke Russian was less important to voters in Ukraine than their policy positions. There is also evidence of renewed civic identity in Ukraine, at least relative to civic identity in Russia (Goble 2016). While Sasse and Lackner (2018) show that the increased feeling of Ukrainian identity is stronger among the Ukrainian-speaking and bilingual population as opposed to monolingual Russian-speaking Ukrainians, Giuliano (2018, 170) documents a "trend of non-mobilization" along ethnic lines persisting after the outbreak of the war, arguing that material interests and responses to national-level political developments were more important than ethnic identity (Giuliano 2015). All of these studies examine explicitly held views, meaning that we still know very little about the implicit associations of these respondents. Precisely because ethnic identity appears to be low-salience in Ukraine, application of the IAT will help us better understand the degree to which explicitly stated attitudes match implicit associations.

Third, all the cities in which we ran the lab have both self-identified ethnic Russian and Ukrainian populations, but their different locations mean they have had different experiences with the conflict (see map in figure 2). Some cities are more affected and more at risk of being affected than others, with Kharkiv at the greatest risk during the study period. Of all the cities in which we collected data, Kharkiv was the closest to the heavy fighting around Debaltseve in February 2015. It was also the scene of intense riots and unsuccessful attempts by anti-Ukraine insurgents to seize government buildings and territory earlier in the conflict. Moreover, the city is closest to Ukraine's long border with Russia, making it most at risk from the buildup of Russian military equipment and personnel along the border that began early in the conflict. Data from a nationally representative survey conducted in May 2015, one month after our data collection, showed that respondents in Kharkiv had the highest expectation of conflict-related violence compared to the other three cities.[3] In short, at the time of data collection, Kharkiv was the city out of our four locations most likely to come under threat from Russia-backed insurgents or Russian military action.[4]

Theoretical Framework

The "rally 'round the flag" effect is a surge in public support for a political leader or government in the wake of a major international crisis or war. These rallies have occurred in countries and conflicts as diverse as the United Kingdom during the Falklands War and Persian Gulf War (Lai and Reiter 2005), the United States following the terrorist attacks of September 11, 2001 (Kam and Ramos 2008), and Russia after it invaded the Ukrainian region of Crimea in 2014 (Theiler 2017; Hale 2018). Cross-national quantitative analysis finds that leaders may initiate international conflict in order to increase their support at home by exploiting this rally effect (Davies 2002).

There are multiple proposed mechanisms for the rally effect, but one of the most studied is the "patriotism" mechanism (Mueller 1973). Patriotism links an individual's reaction to international conflict and their change in opinion about their government, drawing on the logic of social identity theory (Tajfel and Turner 1979). According to social identity theory, individuals derive personal value from group identity, want to have positive feelings toward their group, and develop an in-group bias relative to out-groups. Proponents of the patriotism explanation for the rally effect argue that military conflict creates a newly salient, foreign enemy, a new out-group, casting the entire nation as the relevant in-group. This stronger sense of identification with the entire nation during times of crisis and war leads to a more positive appraisal of the nation and, by extension, the political figures and institutions that represent it. Individuals feel a strong bias in favor of their home country relative to the adversary, and political leaders benefit by way of a surge in popular support.

While generally supporting the rally effect and the patriotism mechanism, prior work has also noted that the effect of the rally is mediated in important ways by political ideology (Merolla and Zechmeister 2013; Kobayashi and Katagiri 2018) and prior support for head of state and government policy (Sigelman and Conover 1981), as well as identity factors like race and gender (Perrin and Smolek 2009). Still others have noted that subpopulations experience very different emotional reactions to events that produce a rally within the population at large (Ojeda 2016). We extend this research by examining how ethnic identity and geography may predict different attitudes toward the home state and the aggressor against the backdrop of a wartime rally caused by a conflict that has potentially important ethnic and spatial dimensions.

Researchers of conflict, civil war, and the comparative study of ethnicity have shown that ethnic divisions can become highly salient in conflict settings, marking the boundaries of relevant in- and out-groups in a way that could cause different ethnic groups within a given state to develop different views on the same conflict. Ethnic divisions and ethnic exclusion can contribute to explaining both the duration (Fearon 2004) and onset of civil wars

[3] For more information on the survey, see the Results section, below.

[4] This grouping is subjective, and a case can be made for considering all of the cities except Kyiv as proximate to the conflict and vulnerable to spillovers in violence, a possibility we consider in the online appendix A.

(Wimmer 2013; Buhaug, Cederman, and Gleditsch 2014), particularly over territory seen as a group's homeland (Toft 2010). Acts of violence, manipulation of information, or perceived injustices by the ethnic other can activate ethnic animosity, even where it did not exist previously (Petersen 2002). At the most extreme, conflict that is fought along ethnic lines "destroys the possibilities for ethnic cooperation" (Kaufmann 1996, 137), making multiethnic society impossible.[5] In these instances, conflict is fundamentally about questions of whether a group is ruled over by its own members, defined in ethnic terms.

Ethnicity can become salient or emerge as an imagined category of social distinction (Anderson 1991), even in societies where individuals historically attached relatively low salience to ethnic identity or often did not think of themselves in such terms, such as many in Ukraine before the start of the conflict in 2014 (Shevel 2002). Kuran (1998) argues that an exogenous shock can cause *ethnification*, whereby individuals increasingly coordinate behavior along ethnic lines in expectation that others will also do so. This reputational cascade can lead some to express greater ethnic attachment than they feel, a type of ethnic preference falsification. This process is similar to the process of "ethnic polarization" identified by Somer (2001, 128), in which an exogenous shock pushes individuals to embrace a "divisive image" of ethnicity, one that "implies a definition of ethnic identities as mutually exclusive and incompatible with belonging to the same nation." In this account, individuals begin to behave in ways consistent with their new conceptualization of polarized and mutually exclusive ethnic categories.

We argue that war between states with different majority ethnic groups could provide an exogenous ethnification shock to those groups. Indeed, the rally effect, which demonizes the adversary state, could create or increase feelings of persecution, insecurity, and fear among the ethnic minority with ties to the aggressor. Under such conditions, ethnic identity would become more salient for ethnic minorities who are co-ethnics with the majority group in the adversary state.[6] Even if they had no feelings of antipathy toward the host state previously, ethnic minorities could come to fear that state during a conflict with their external sponsor because states in low-information environments use ethnicity as a marker for loyalty, repressing those groups they perceive as disloyal (Wallace 2015; Blaydes 2018; McNamee and Zhang 2019). Such a scenario could cause ethnic identity to become the salient group-level identity *within each state in the conflict*. Under such circumstances, ethnic identity would predict attitudes toward the warring states, with the non-core ethnic group in each country sympathizing with its external supporter, i.e., the adversary of the home state, as stated in *Hypothesis 1*.

Hypothesis 1. *Ethnic identity will predict bias, with the dominant ethnic group expressing bias for the home state and the minority ethnic group expressing bias in favor of the external adversary with which it has ties.*

If *Hypothesis 1* is correct, then we would expect citizens of Ukraine who identify as ethnic Russians to express pro-Russia attitudes, while those who identify as ethnic Ukrainians would be pro-Ukraine.

The logic that leads to *Hypothesis 1* above assumes that the conflict will fracture society along ethnic lines, but this need not be the case. War with a foreign enemy can minimize the role of ethnic identity relative to state identity, producing a nationalism that is based on the state and is necessarily more civic in nature than ethnic.[7] Indeed, drawing on the European political development literature Herbst (1990, 122) argues, "the threat of a palpable external threat may be the strongest way to generate a common association between the state and the population."

In such a situation, all ethnic groups express a bias in favor of the home state, but ethnic minorities who are co-ethnics with the majority group in the adversary state may feel less antipathy toward the adversary than individuals of other ethnic groups. The relatively less pronounced feelings of antipathy toward the aggressor would result in a less pronounced pro-home-state bias. Thus,

Hypothesis 2. *Ethnic identity will predict variation in the degree of bias in favor of the home state.*

If *Hypothesis 2* is correct, then we would expect both ethnic Russian and ethnic Ukrainian residents of Ukraine to express pro-Ukraine attitudes, although at different levels; those who identify as ethnic Russians would have a weaker bias in favor of Ukraine over Russia.

Contrary to group-level theories of identity, others have focused on physical proximity to and experience with wartime violence as the key factor in explaining a range of attitudes and behaviors during and following conflict. Gibler, Hutchison, and Miller (2012) suggest that individuals in the state that is attacked will identify more with that state as they need its protection, but that this effect is stronger closer to the conflict zone. Similarly, Tellez (2019) shows how those living closer to a conflict zone express different attitudes toward the conflict than those who are farther away. While a different kind of international security situation, Cortina (2020) shows that attitudes toward a highly politicized issue—the proposed border wall between the United States and Mexico—are a function of distance from the border.

If proximity to conflict can affect identification with the state, attitudes about the conflict, and policy preferences, then we would expect individuals in areas closer to fighting to express attitudes about the belligerents that are different from those in areas farther away. There are many ways that this could play out. Individuals who are closer to the conflict could overstate their loyalty to the home state in an attempt to demonstrate their allegiance and win protection from possible attack. On the other hand, if they foresee being caught up in the conflict, individuals may hedge their responses by downplaying their preference for one state over the other, trying to stake out a neutral position. Alternatively, if they view a takeover by the adversary state as likely they may express a preference for the adversary over the home state. While the exact nature of the difference is difficult to predict, what is consistent across all these scenarios is that those who are closer to the fighting are likely to express views toward the home state and the belligerent that are at odds with respondents who are more removed from the conflict:

[5] Admittedly, many downplay the role ethnic diversity plays as an *independent* cause of civil war (Collier and Hoeffler 2004).

[6] This situation would be similar to the one described by Mylonas (2012, 26–32) of a "non-core group" in a state that is in conflict with the "external power" that supports it.

[7] For instance, Saideman and Ayres (2008) demonstrate how the external threat of Soviet invasion in 1956 resulted in increased pro-Hungary feelings, which helped unify the multiethnic state and inculcate a strong new Hungarian identity.

Hypothesis 3. *Greater proximity to the conflict will predict different levels of bias toward the home state and the aggressor.*

Coming up with an a priori distance within which attitudes will be swayed by risk of exposure to conflict is difficult. Bakke, O'Loughlin, and Ward (2009), for instance, use a 50-km threshold in their operationalization of proximity to conflict in the North Caucasus, but that was due to an analysis of the geography of violence in that case. As we have only four cites in which we ran the IAT, taking a similar approach and finding the appropriate cutoff on a continuous distance measure is not feasible. Instead, looking at the four cities in which we ran the IAT, our *Hypothesis 3* expects that respondents in Kharkiv would hold the most conflict-affected attitudes in their views of Ukraine and Russia as compared to respondents of the three other cities (Kyiv, Kherson, and Odesa). We believe that, at the time that the study was in the field, Kharkiv was at greater risk than the other cities of becoming embroiled in the conflict due to its geographic proximity to active or potential conflict and the violence that had already occurred within the city. We acknowledge that proximity to the conflict is not the only dimension along which Kharkiv is different from the other cities, making it an imperfect measurement. To address this problem, as well as the subjective nature of these types of groupings, we also consider an alternative operationalization of *Hypothesis 3*, in which Kharkiv, Kherson, and Odesa are considered more proximate to the conflict, and attitudes in those cities are contrasted to those in Kyiv (see the online appendix A).

Implicit Association and Explicit Attitudes

Traditional surveys evaluate respondents' explicit attitudes, which are the product of a psychological process that includes the interaction of both affect and cognition. As a result, it can be difficult to tell whether a respondent's answer to a question is the result of a process of reasoning and deliberation by which they arrive at an answer (cognition) or a strong underlying feeling (affect). These underlying feelings—implicit associations—are worth studying and understanding because they have been shown to predict important political behavior and attitudes, including party choice and turnout (Arcuri et al. 2008; Ryan 2017), attitudes toward female politicians (Beaman et al. 2009; Mo 2015), and immigration policy preferences (Pérez 2010), to name just a few. In some cases, implicit attitudes are more accurate predictors of behavior than explicit attitudes (Roccato and Zogmaister 2010). Studying implicit attitudes is particularly important in sensitive situations or when seeking to understand how individuals relate to taboo topics because it is difficult for respondents to fake their answers—social desirability bias and other forms of preference falsification do not affect implicit associations.

We follow Truex and Tavana (2019), who studied attitudes toward political leaders in an authoritarian setting, by comparing implicit and explicit biases to understand public opinion about an ongoing military conflict. We use the associative-propositional evaluation (APE) model (Gawronski and Bodenhausen 2006) of attitude formation to understand the progression from implicit associations to explicit attitudes. According to this model, implicit associations are automatic responses to a stimulus. Whether this response is negative or positive is a function of the associations that the stimulus generates within a respondent. In forming explicit attitudes, respondents evaluate their implicit associations in light of propositions, i.e., statements that have truth-value for each respondent. If the propositions lead to an explicit attitude that is consistent with the association, e.g., both are positive, then there is *congruence* between the implicit and the explicit. Previous work has demonstrated that we should generally expect to find congruence (Nosek, Banaji, and Greenwald 2002), but the process of evaluating the association with propositional reasoning may lead some respondents to state an explicit attitude that is at odds with their underlying, implicit association, a phenomenon referred to as *dissociation*.[8]

The examination of respondents' implicit bias toward Ukraine or Russia will reflect their automatic, underlying feelings, while comparing their explicit attitudes toward the two countries will filter those implicit association through a process of cognitive reasoning. As there are two possible outcomes each for implicit associations and explicit attitudes, there are a total of four possible outcomes:

- **Pro-Ukraine congruence**: The respondent is implicitly and explicitly pro-Ukraine.
- **Pro-Russia congruence**: The respondent is implicitly and explicitly pro-Russia.
- **Pro-Ukraine dissociation**: The respondent is implicitly pro-Russia but is explicitly pro-Ukraine or neutral. This type of dissociation is what one would most expect to find if the conflict setting is leading people to engage in a cognitive process that reverses their implicitly pro-Russia orientation.
- **Pro-Russia dissociation**: The respondent is implicitly pro-Ukraine but is explicitly pro-Russia or neutral. We expect that this will be quite rare in light of the fact that the ongoing conflict should encourage and support favorable assessment of Ukraine, but not of Russia.

It is important to keep in mind that dissociation is not necessarily the same thing as social desirability bias, in which respondents misrepresent their *explicit* attitudes or behaviors to avoid reprisal or some social sanction. When preference falsification may be a problem,[9] scholars have employed sensitive survey methods such as list experiments or randomized response to remove the effect of social desirability bias on understanding corrupt practices (Malesky, Gueorguiev, and Jensen 2015; Brierley 2020), support for the leader in authoritarian regimes (Frye 2015), or attitudes toward militant groups in conflict settings (Lyall, Blair, and Imai 2013; Blair, Imai, and Lyall 2014). These techniques allow respondents to state an *explicit* view that they would otherwise not feel comfortable sharing. While these methods allow researchers greater confidence that they have elicited a truthful expression of an explicit view, they provide no information about how respondents may feel about a given subject on an automatic or reflexive level, i.e., their implicit association toward a given subject. These methods cannot detect dissociation—the divergence of implicit association and explicit bias—which may still occur even if respondents are truthfully stating their explicit attitudes.

Research Design

After conducting a pilot in January of 2015 in Kharkiv, we ran the IATs in April 2015 in four cities: Kyiv, Kharkiv, Kherson, and Odesa (See figure 2). These IATs, as well as nationally representative survey data used in robustness

[8] Neither implicit associations nor explicit attitudes should be thought of as necessarily more fixed or as true beliefs, and changes in the relevance of associations or propositions can cause shifts in both implicit and explicit biases (Gawronski and Bodenhausen 2006).

[9] Kuran (1991) provides a theoretical account of this process.

checks, were part of a research effort that was commissioned and funded by the National Democratic Institute (NDI). The Kyiv International Institute of Sociology (KIIS) carried out the data collection.

Subject Recruitment

In each of the four cities, we recruited 150 subjects to take part in our lab-in-the-field from pre-specified sampling locations throughout each city. All subjects were told that they would be given a test about their attitudes and we reported the implicit score to them at the end of the test. A pre-test had shown that we were able to recruit a sufficiently wide range of individuals who self-identified with Russian or Ukrainian ethnicity, but we were concerned about our ability to recruit a sufficient number of individuals from across the political spectrum (for descriptive statistics on our respondents by city, see the online appendix B). To address this potential problem, we used quotas to allocate the sample within each city, recruiting on reported voting behavior from the 2014 parliamentary election (abstainers, pro-Russia voters,[10] and anti-Russia voters), age, and gender. Allocating quotas in this way only omitted voters who supported far-right parties in the 2014 elections, a small fraction of the adult population in our cities of interest.

Within each city, we recruited approximately fifty subjects who had voted for pro-Russia parties, fifty who had voted for anti-Russia parties, and fifty who had abstained. Our rationale for this recruitment strategy was that pro-Russia voters might be more likely to explicitly or implicitly favor Russia and, in the case of having pro-Russia implicit attitudes, would be most likely to dissociate. Including abstainers was particularly important because those who did not vote could be systematically different from those who did. While quota sampling can introduce bias, no laboratory sample is representative, and we wanted the ability to analyze different patterns from different sample populations.[11]

Measuring Implicit and Explicit Attitudes

The IAT has a long track record of uncovering respondents' implicit biases and it has several qualities that aid researchers in studying these biases. The IAT has strong test-retest reliability (Lane et al. 2007); it is hard for respondents to manipulate (Greenwald, Nosek, and Banaji 2003; Greenwald et al. 2009); and its measures of implicit attitudes have been shown in multiple contexts to be strong predictors of political behavior (Arcuri et al. 2008; Pérez 2010; Roccato and Zogmaister 2010). Following Greenwald, Nosek, and Banaji (2003), we measure IAT respondents' explicit bias as the difference ($\boldsymbol{Ukraine}_i - \boldsymbol{Russia}_i$) between their stated responses to two seven-point thermometer questions, which asked about their attitudes toward Ukraine and Russia.[12] The assumption in measuring differences is that the difference in preference is what matters, not the overall level of the preference. Therefore, a respondent who rated Russia 1 and Ukraine 2 would have a pro-Ukraine explicit preference score of 1, as would a respondent who ranked Russia 5 and Ukraine 6.

IATs were used to measure respondents' implicit bias toward either Ukraine or Russia.[13] These computer-based tests prompt respondents to associate words with a given *category* (Ukraine or Russia in our case) or a given *attribute* (Positive or Negative in our case). In the test, a word that is associated either with a category or an attribute is displayed in the middle of the screen, while the corresponding categories or attributes are displayed in the upper left and upper right corners. Using the "E" and "I" keys on the keyboard, the respondent is asked to associate the word in the middle of the screen with the relevant category or attribute as quickly as possible.[14] The computer tracks the time that it takes a respondent to perform each such association task, generating a metric referred to as the *response latency*.

The validity of the IAT comes from the fact that if a respondent does not associate a category (e.g., Russia) with the attribute (e.g., Negative) listed in the same side of the screen, then the respondent will be much slower in choosing the side of the screen to which the word in the middle belongs. For example, respondents will be slower to associate positive words (such as "love") from the middle of the screen when a country they dislike is paired with the Positive category. Each respondent's implicit bias toward either Ukraine or Russia is the standardized difference (IAT *d*-score) between that respondent's response latencies on blocks where the negative attribute is paired with Ukraine and positive attribute with Russia, and blocks where the negative attribute is paired with Russia and positive attribute with Ukraine.[15]

Our IAT proceeded as follows. First, respondents completed training rounds on words that are associated with our two categories: Ukraine and Russia. Next, they completed training rounds on words that are associated with our two attributes: positive and negative. After the training rounds, respondents proceeded to the main blocks of the test. In a first set of blocks, each of the two categories (Ukraine and Russia) were shown to respondents on either the upper left or right hand corners of the screen, next to one of the two attributes (Negative or Positive). In a second set of blocks, the category associated with the attribute is reversed: if Negative was shown next to Russia and Positive next to Ukraine in the first set of blocks, then Negative was shown next to Ukraine and Positive next to Russia in the next set of blocks. Figure 3 shows a sample screen.

Respondents took the IAT in Russian because we wanted to standardize the language in which subjects took the test. However, we acknowledge that this is a potential source of bias because it could exclude some people from sampling if they do not read Russian. We discuss this possible source of bias in our discussion of robustness in the results section.

We construct both explicit and implicit scales such that values greater than zero imply preference for Ukraine, while values less than zero imply preference for Russia. We

[10] See online appendix C for party classification. Abdelal (2005, 116) describes the parties these voters support as "not anti-Russian." We use the term "pro-Russia" because these parties are more in favor than other parties of working to normalize relations and deepen ties with Russia. Voters may support these parties for many other reasons besides their stance toward Russia, such as Soviet nostalgia or support of individual local politicians. Giuliano (2018) and Toal (2017) further discuss the motivations of "pro-Russia" party voters in the context of the ongoing conflict in eastern Ukraine.

[11] We opted for a lab environment instead of online samples because the lab setting allowed us to explain and debrief to subjects the study in more detail, which we deemed important. The lab environment also delivered near perfect compliance with the IAT protocol.

[12] The questions read: "How do you feel about the following countries?" Respondents were then asked about several countries including Ukraine and Russia. The scale ranged from "Strongly negatively" (−3) to "Strongly positively" (3).

[13] The IAT software *Implicit Millisecond* was used to carry out the tests.

[14] The computer keyboard used in the lab had both English and Cyrillic letters pasted on the keys.

[15] The online appendix D explains how the *d*-score is calculated and lists the words used in the IAT.

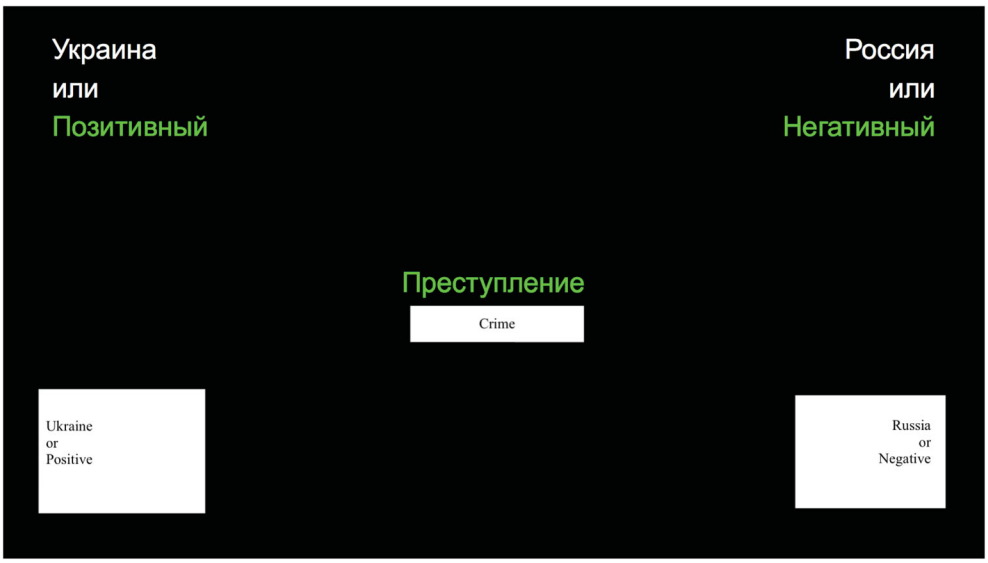

Figure 3. IAT screen example. In all blocks with both categories and attributes, categories are displayed in white at the very top of the screen, with attributes immediately beneath them (the test computer displayed attribute text in green). In this example, the Ukraine category is paired with the "Positive" attribute on the left and the "Russia" category is paired with "Negative" on the right. The word that the respondent must match with the corresponding category or attribute appears in the middle of the screen. In this example, it is a negative attribute ("Crime"), so it appears in green. To answer correctly, respondents need to associate "Crime" with "Russia or Negative." While not present in the actual test, English translations are shown for this example in white boxes.

Table 1. Summary statistics of explicit and implicit preferences by ethnicity, with values greater than zero indicating pro-Ukraine bias

	Pro-Ukraine explicit					*Pro-Ukraine implicit*				
Ethnicity	*Mean*	*SD*	*Minimum*	*Maximum*	*n*	*Mean*	*SD*	*Minimum*	*Maximum*	*n*
Ukrainian only	2.42*	2.22	−5.00	6.00	386	0.53*	0.39	−0.73	1.64	386
Other	1.63*	2.53	−5.00	6.00	54	0.28*	0.45	−0.84	1.05	54
Ukrainian and Russian	0.91*	2.36	−6.00	6.00	80	0.26*	0.45	−0.87	1.32	80
Russian only	0.12	2.43	−6.00	6.00	80	0.12	0.56	−1.15	1.33	80
All groups	1.84*	2.44	−6.00	6.00	600	0.42*	0.46	−1.15	1.64	600

Notes: Means of the explicit measure are of the full range of possible values (i.e., integers in [−6, 6]), while the means of the implicit measure are calculated from the *d*-score units (i.e., real values in [−1.15, 1.64]). *indicates mean is distinct from 0 with 95 percent confidence.

report summary statistics (see table 1) for the explicit scores in the units of the difference in Likert scales (integers in [−6, 6]) and for the implicit bias in *d*-scores, a standardized measurement of implicit bias. In both cases, greater than zero implies a pro-Ukraine bias, but the scales are not directly comparable to each other.

Operationalizing Predictors and Other Control Variables

To test the role of *ethnicity* (Hypothesis 1 and Hypothesis 2), we ask respondents to list the ethnic group (*ėtnicheskaīa gruppa*) or groups to which they feel they belong, allowing respondents to list multiple ethnic groups.[16] We use two measures of ethnicity in our analysis. Our first approach groups anyone who lists Russian as a response in the Russian category and places all others in the non-Russian category. Our second approach, and the model specification results on which we focus, separates out four groups: those who identified as only Ukrainian, those who identified as only Russian, those who identified as both Russian and Ukrainian, and all other respondents.[17] By holding the lab in four different cities, our research design allows for a test of *proximity to conflict* (Hypothesis 3) by including controls for city in the statistical model. Due to findings that ethnicity can have differential effects in different regions of Ukraine (Erlich and Garner 2020), we interact our city variable with our four-category measure of ethnicity in our main model specification.

Covariates other than ethnic identity and proximity to conflict may also predict implicit or explicit bias. To address this, we control for economic, social, and political variables that may be correlated with our outcomes of interest. We include controls for whether respondents are *unemployed*, because unemployment could drive less connectedness or greater dissatisfaction with Ukraine; and whether a respondent has *family in Russia*. We also include controls for *age*, *gender*, and whether the respondent's *home language*

[16] The exact wording of the open-ended question was: "Many people identify with one or more ethnic groups. Please state the ethnic group or groups that you identify with."

[17] This second measurement strategy emulates the new approach taken by the US census and allows us to determine whether those who identify with overlapping identities are different from those who do not, mirroring a concern of Onuch and Hale (2018, 91) in Ukraine.

was Russian or Ukrainian, as Onuch and Hale (2018) and Kulyk (2011) show that reported ethnic identity is not the only relevant or salient component of ethnicity in Ukraine, with language playing a meaningful and distinct role. For the model of explicit bias, we also control for *voting behavior* in the 2014 parliamentary election (on which we recruited and which is an important residual category of sentiment).

In our statistical models of implicit and explicit bias, we regress these outcome variables on the predictors and control variables using OLS regression with robust standard errors, although there are two important differences between the models. First, our model of explicit bias controls for respondents' implicit association scores, but the opposite is not true because in the APE model the explicit attitude is formed only after the implicit association has been made. Second, and for largely the same reason, we do not include voting behavior in the model that predicts implicit bias because we believe voting behavior is a product of implicit bias, not the other way around.[18]

Results

In our combined sample, the average respondent expressed a pro-Ukraine explicit attitude and a pro-Ukraine implicit bias. Moreover, in each of the four ethnic identity categories, the average respondent was pro-Ukraine on both explicit and implicit measures (see table 1). That said, the mean values for explicit and implicit bias among those who identify as ethnic Russian are not greater than zero at standard levels of statistical significance, indicating that ethnic Russian identity may be an important predictor of relative levels of bias. There were differences in the mean levels of bias from city to city, too, although all were on average pro-Ukraine (see tables B2 and B3 in the online appendix). The explicit bias measure ranged from an average of +1.4 in Odesa to +2.6 in Kyiv. This variation was driven mostly by assessment of Russia, as explicit attitudes toward Ukraine ranged from an average of 5.4 in Kharkhiv to 5.7 in Kyiv, whereas differences in average explicit attitudes toward Russia were wider, ranging from 3.2 in Kyiv to 4.1 in Odesa. These initial summary statistics run counter to Hypothesis 1, but they provide suggestive evidence to support Hypothesis 2 that, in the context of a wartime rally, ethnic identity predicts relative levels of preference for the home state over the adversary. They also suggest that, as predicted by Hypothesis 3, proximity to conflict can predict bias.

Regression Analysis

Our main empirical models provide evidence that is largely consistent with the findings above (see models 5 and 6 in table E1 in the online appendix) The top-left and top-right panels of figure 4 present the model-based predicted values for those respondents who self-identify as ethnic Russians only (see figure E1 in the online appendix for a complete set of predicted probabilities by ethnicity and city). After controlling for relevant covariates, those who identify as ethnic Russians are consistently pro-Ukraine (both explicitly and implicitly), although the predictions are not always statistically distinct from zero at the 95 percent confidence level. This finding is a clear rejection of Hypothesis 1; on average, ethnic Russians neither explicitly nor implicitly favor Russia over Ukraine. While we are unable to say that pro-Ukraine views were caused by the conflict in the east, we can say that in the context of a wartime rally, those who identify with the ethnic minority with ties to the aggressor are still, on average, biased in favor of the home state on both an implicit and explicit level.

The bottom two panels of figure 4 show the first difference associated with a change in ethnic identity from Ukrainian to Russian to test whether ethnic identity may predict meaningful differences in pro-home state bias, on average, per Hypothesis 2. In all cities, and for both explicit and implicit biases, modeling a change from ethnic Ukrainian to ethnic Russian identity leads to less of a pro-Ukraine bias, although not always at statistically significant levels. Thus, even after controlling for theoretically relevant factors, it appears that ethnic Russian identity is associated with a weaker bias in support of the home state over the adversary.[19] Taken together, the model-based predicted values and first differences for ethnic Russian identity provide clear support for *Hypothesis 2*. Ethnic Russian respondents are not pro-Russian, but they are less pro-Ukraine than ethnic Ukrainians.

Finally, figure 4 allows us to examine whether Kharkiv, the city most proximate to the conflict, is systematically different from those cities farther away, per Hypothesis 3. While there are differences between the cities, no clear pattern emerges. In Kyiv and Kherson there is a large and statistically significant difference between the explicit biases of ethnic Ukrainians and ethnic Russians, with no statistically significant difference between these groups in Odesa or Kharkiv. However, the pattern does not hold for implicit bias. In Kyiv, there is no statistically significant difference in implicit bias between those who identify as Ukrainian and those who identify as Russian, while in Odesa, Kherson, and Kharkiv there is a larger and statistically significant difference. These findings show no consistent difference between Kharkiv and the other cities, and thus no clear association between proximity to conflict and whether respondents are more or less pro-Ukraine in their explicit or implicit biases.

Dissociation in Implicit and Explicit Attitudes

Comparing individual respondents' implicit and explicit biases reveals that some respondents are stating explicit attitudes that differ from their implicit biases. Pooling all respondents, explicit and implicit attitudes were correlated at $\rho = 0.33$ (see figure 5), a level that indicates a reasonably high degree of attitude dissociation. Slightly more than one in three respondents (37 percent) dissociated. To put these figures in perspective, we compare them to the correlation of explicit and implicit attitudes in a very permissive context and a relatively repressive context. Explicit and implicit attitudes toward US President George W. Bush during the 2004 presidential race were correlated at $\rho = 0.73$ (Greenwald, Nosek, and Sriram 2006), while research conducted in 2016 showed Egyptians' attitudes toward President Abdel Fattah El-Sisi were correlated at $\rho = 0.17$ (Truex and Tavana 2019). The level of dissociation we find in our study indicates that something in some respondents' process of propositional reasoning is leading to an explicitly stated preference that is at odds with the implicit association. This could be driven by internal processes or by social desirability bias or other forms of preference falsification.

Pro-Ukraine dissociation was much more common than pro-Russia dissociation, consistent with our expectation that

[18] Results are substantively unchanged in alternate specifications that include voting behavior in the model that predicts implicit bias.

[19] As a robustness check, we conduct a coarsened exact matching (CEM) exercise to verify that there is a relationship between Russian ethnicity and lower levels of pro-Ukraine bias.

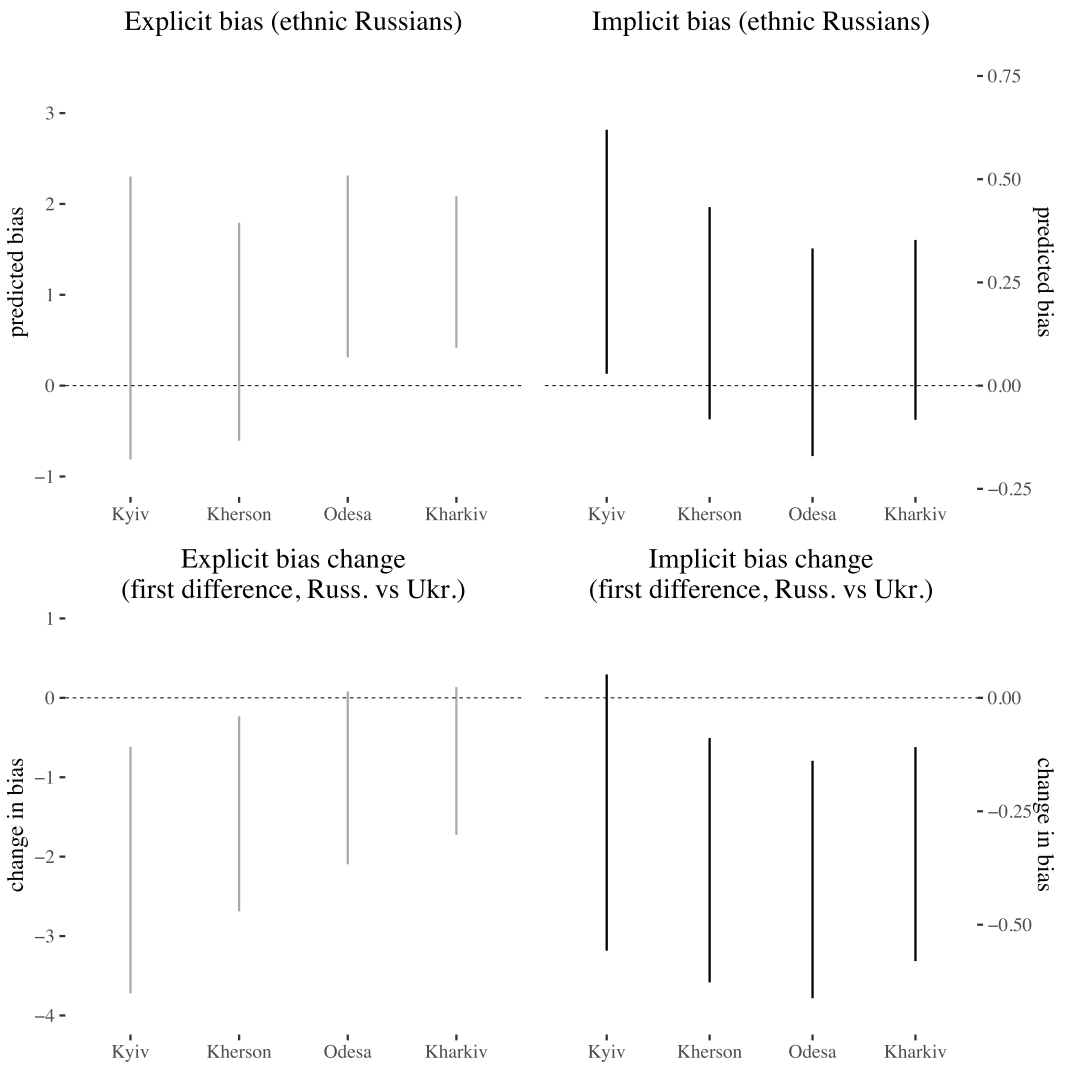

Figure 4. Predicted levels of bias and first differences from main statistical models (see table E1 in the online appendix). Implicit bias is measured in *d*-score units, while the predicted values for explicit bias reflect the range of possible values in [−6, 6].

dissociating in favor of the home state would be more common during war. Out of all implicitly pro-Russia respondents ($n = 100$), 70 percent reported a pro-Ukraine or neutral explicit attitude, while among implicitly pro-Ukraine respondents ($n = 500$) just 30 percent stated a neutral or pro-Russia explicit view. This finding suggests that a conflict environment can have a differential effect on the rates of dissociation between categories. While pro-aggressor dissociation was much less common, we were surprised that so many respondents dissociated *in favor* of Russia, with 150 individuals stating neutral or explicitly pro-Russia views despite having an implicit bias for Ukraine. These individuals included respondents from all ethnic groups, including 22 percent of those who identified as only Ukrainian and 32 percent of those who identified as only Russian, and in all cities.

Finally, an analysis of those identifying as ethnic Russian belies the notion that ethnic minorities are likely to hide pro-aggressor views. Indeed, more ethnic Russians reported a pro-Russia explicit view despite being implicitly pro-Ukraine than expressed explicit bias for Ukraine when they implicitly favor Russia. Ethnic Russians were least likely, when compared with other ethnic groups, to state a pro-Ukraine explicit view if they were implicitly pro-Russia. This finding contradicts the classical expectations about how dissociation would work in an ethnic minority group with ties to the adversary and challenges the notion that ethnic minorities are secretly part of a so-called fifth column. After controlling for theoretically relevant covariates, the relationship between Russian ethnic identity and dissociation is inconsistent across our different model specifications, with no statistically significant relationship in our fully specified model. The coefficient associated with being implicitly pro-Russia is always large and significant, regardless of model specification (see table E2 in the online appendix). Further, after controlling for theoretically relevant covariates, dissociation appears more likely to occur in Kharkiv, but the relationship is never statistically significant at standard levels across our various model specifications.

Additional Findings

In addition to the findings regarding our main hypotheses, our model reveals three additional important findings. First, those who identify as only ethnic Ukrainian express the strongest pro-Ukraine implicit and explicit biases. After

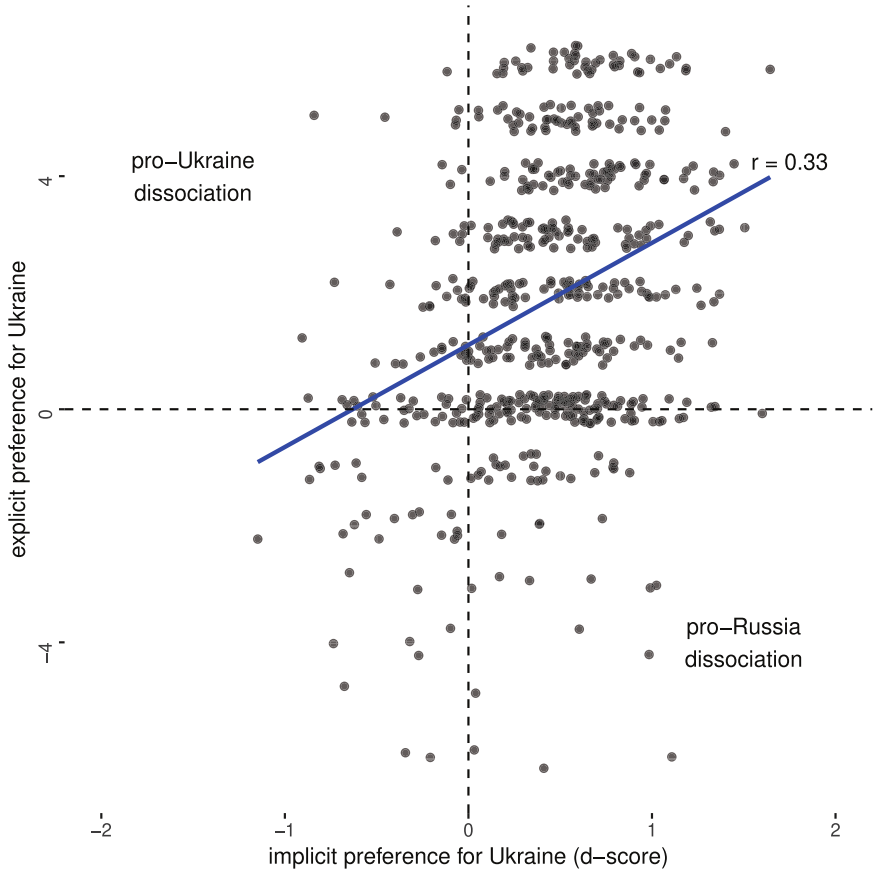

Figure 5. The majority of IAT participants expressed pro-Ukraine implicit and explicit preferences (upper right).

controlling for theoretically relevant covariates, predicted values for ethnic Ukrainians are consistently more pro-Ukraine for both explicit and implicit biases in all four cities (see figure E1 in the online appendix). Second, those who claim both Russian and Ukrainian ethnic identities are more similar in their explicit and implicit biases to those respondents who claim only Russian ethnicity than to those who claim only Ukrainian identity, particularly in Kyiv. Third, there was an unexpected finding regarding having voted for a pro-Russia political party. Supporters of pro-Russia parties expressed an explicit bias for Ukraine over Russia, although they were less explicitly pro-Ukraine than those who voted for anti-Russia parties.

Robustness

The quota sampling strategy is one possible vector by which we may inadvertently bias our results. To address this possibility, we conduct two analyses. First, we reweight the data on our recruitment categories (pro-Russia voting, anti-Russia voting, and abstaining) using city-level data from Ukraine's Central Election Commission on the outcome of the 2014 parliamentary elections. Specifically, we weight the pro-Russia vote and anti-Russia vote observations in our data to the number of votes cast in each city for pro-Russia and anti-Russia parties on the national party list component of the ballot. We weight the abstainers to the difference between votes cast and eligible voters in each city. After weighting the data, the predicted probabilities of interest are virtually unchanged (see figure F1 in the online appendix).[20] Additionally, the level of dissociation that we see is only slightly different after weighting the data, falling from $\rho = 0.33$ to $\rho = 0.31$ (see figure F2 in the online appendix), and we find no meaningful differences in the significance or signs on our coefficients of interest (see table F1 in the online appendix).

Second, we compare the explicit evaluation of Russia that we collected from respondents in the lab to data from a nationally representative survey. Data from this survey show that the majority of respondents in each city where we fielded the IAT and in each ethnic group believe Russia has a negative or very negative influence on Ukraine.[21] This response suggests that the population we recruited, which gave Russia roughly equivalent scores on the explicit questionnaires, is not vastly different on our explicit outcome of interest than the populations of each city in which we fielded the IAT.

The language of the IAT (Russian) is another possible vector by which bias could have been introduced into our results by essentially requiring that potential respondents speak Russian, but even if there was an effect, we believe

[20] The sign on the predicted bias for ethnic Russians in Odesa switches from positive to negative, but the confidence interval overlaps zero.

[21] The question was: "In your opinion is the influence of the following, in determining Ukraine's future, very negative, negative, neither negative nor positive, positive, or very positive?" The face-to-face survey of 5,847 residents of Ukraine carried out by KIIS in May 2015 included oversamples ($n = 600$) in three of the four cities in which we ran our lab: Kharkiv, Kyiv, and Odesa. Kherson had fewer observations ($n = 191$), leading to greater uncertainty, but we believe this sample provides a reasonable comparison.

it would be quite small. In Ukraine, particularly in 2015 in the cities in which we fielded the IAT, we believe very few potential respondents would have been excluded from recruitment by this requirement. Ukrainian society is largely bilingual, and many people prefer to speak Russian in their daily lives. Data from a May 2015 survey, just one month after our data collection, corroborate this, as the vast majority of randomly selected respondents in the cities where we ran our labs answered the survey in Russian rather than in Ukrainian (99.3 percent in Kharkiv, 100 percent in Kherson, 84.8 percent in Kyiv, and 96 percent in Odesa).[22] Still, we cannot eliminate the possibility that our results are biased by the exclusion of those who do not speak Russian or choose not to do so. However, we argue that if this were the case it would lead us to underestimate the amount of pro-Ukraine sentiment, if anything, as we believe that not knowing Russian or refusing to speak Russian would be associated with stronger pro-Ukraine feeling.

Conclusions

The wartime "rally ′round the flag" effect has occurred in many countries around the world and across many types of conflict, from conventional wars to terrorist violence. Our research contributes to understanding public opinion against the backdrop of a rally in two ways. First, we show that ethnic minority populations are not necessarily favorable to the country with which their ethnicity is associated. Second, we find that ethnicity can be an important predictor of the intensity of bias in favor of the home state for ethnic minority populations with ties to the aggressor state.

Taken together, these findings make an important contribution to the literature on ethnicity, identity, and conflict. In light of prior work showing that conflict involving multiethnic societies can exacerbate ethnic divisions and stoke irredentism, the pro-Ukraine implicit and explicit biases of ethnic Russians in Ukraine urge caution toward assumptions that ethnic minority populations will form so-called fifth columns. We interpret our findings as being consistent with work that demonstrates how identity can cohere in a country that has been the target of aggression, reinforcing a supraethnic national identity rather than fracturing society along ethnic lines. This interpretation is consistent with other findings that have shown renewed strength of civic identity in Ukraine in the wake of the conflict.

Methodologically, we show how an innovative application of the IAT can further both our understanding of attitudes and our confidence in the responses that we obtain through traditional surveys. We find that most respondents have congruent implicit and explicit biases, but dissociation—the divergence of implicit and explicit biases—is not uncommon, with implicit bias for the aggressor being its strongest predictor. Against the backdrop of a conflict with Russia, respondents who are implicitly pro-Russia are much more likely to state explicit biases that are at odds with their implicit biases, a finding that holds after controlling for theoretically relevant covariates. Ethnic identity and proximity to the conflict have a weaker and less consistent relationship to dissociation. The weak relationship between ethnicity and dissociation further undermines the stereotypical view of ethnic minority groups with ties to the aggressor as dissembling or untrustworthy members of a so-called fifth column.

Beyond adding to what we know, our findings raise important questions that urge further study. First, more study is needed to determine a clearer relationship between proximity to conflict and attitudes about conflict. Second, what accounts for the dissociation that we see in our sample deserves further scrutiny. Follow-up studies using sensitive survey techniques could determine whether dissociation is the result of social desirability bias or other external factors. Third, exploring whether having an implicit bias for another state over one's home state predicts dissociation across contexts merits further investigation. Replicating our study in a country that is not at war with its neighbor could help determine the degree to which this finding is associated with interstate conflict, in particular. Fourth, the mechanisms which determine why some ethnic Russians—and some ethnic Ukrainians—exhibit pro-aggressor bias could also be explored further. Possible explanations could be differences in education, media consumption habits, or whether the respondent migrated from Russia to Ukraine. Fifth, whether implicit biases for the home state persist over time is another valuable question. Data on the rally effect show that, in most cases, the effect is temporary, but there has been no research that we are aware of on how implicit bias changes over time in a society that has undergone a wartime rally. Pursuing these questions will further deepen our understanding of public opinion during wartime, along with the interplay of identity and attitudes toward conflict.

Supplementary Information

Supplementary information is available at the *International Studies Quarterly* data archive.

Conflict of Interests

The data for this article were gathered as part of a larger project by NDI, a nongovernmental, nonprofit organization. Both authors have worked as consultants for NDI, but neither believes that this relationship poses a conflict of interest. NDI has no control over the content of the authors' published academic research. The analysis and conclusions contained here solely represent those of the authors.

References

Abdelal, Rawi. 2005. *National Purpose in the World Economy: Post-Soviet States in Comparative Perspective*. Ithaca, NY: Cornell University Press.

Anderson, Benedict R. 1991. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.

Arcuri, Luciano, Luigi Castelli, Silvia Galdi, Cristina Zogmaister, and Alessandro Amadori. 2008. "Predicting the Vote: Implicit Attitudes as Predictors of the Future Behavior of Decided and Undecided Voters." *Political Psychology* 29 (3): 369–87.

Bakke, Kristin M., John O'Loughlin, and Michael D. Ward. 2009. "Reconciliation in Conflict-Affected Societies: Multilevel Modeling of Individual and Contextual Factors in the North Caucasus of Russia." *Annals of the Association of American Geographers* 99 (5): 1012–21.

Balcells, Laia. 2012. "The Consequences of Victimization on Political Identities: Evidence from Spain." *Politics & Society* 40 (3): 311–47.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics* 124 (4): 1497–1540.

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

Blaydes, Lisa. 2018. *State of Repression: Iraq under Saddam Hussein*. Princeton, NJ: Princeton University Press.

[22] Our survey measures actual language usage during the survey, rather than stated preference, which likely explains the high levels of Russian.

Brierley, Sarah. 2020. "Unprincipled Principals: Co-opted Bureaucrats and Corruption in Ghana." *American Journal of Political Science* 64 (2): 209–22.

Buhaug, Halvard, Lars-Erik Cederman, and Kristian Skrede Gleditsch. 2014. "Square Pegs in Round Holes: Inequalities, Grievances, and Civil War." *International Studies Quarterly* 58 (2): 418–31.

Collier, Paul, and Anke Hoeffler. 2004. "Greed and Grievance in Civil War." *Oxford Economic Papers* 56 (4): 563–95.

Cortina, Jeronimo. 2020. "From a Distance: Geographic Proximity, Partisanship, and Public Attitudes toward the U.S.–Mexico Border Wall." *Political Research Quarterly* 73 (3): 740–54.

D'Anieri, Paul. 2007. *Understanding Ukrainian Politics: Power, Politics, and Institutional Design*. Armonk, NY: M. E. Sharpe.

Davies, Graeme A.M. 2002. "Domestic Strife and the Initiation of International Conflicts: A Directed Dyad Analysis, 1950–1982." *Journal of Conflict Resolution* 46 (5): 672–92.

Erlich, Aaron, and Calvin Garner. 2020. "Interpersonal Incomparability in Citizens' Views of Democracy: Survey Evidence from Ukraine." *Journal of Elections, Public Opinion and Parties* 30 (2): 139–60.

Fearon, James D. 2004. "Why Do Some Civil Wars Last So Much Longer Than Others?" *Journal of Peace Research* 41 (3): 275–301.

Fournier, Anna. 2002. "Mapping Identities: Russian Resistance to Linguistic Ukrainisation in Central and Eastern Ukraine." *Europe–Asia Studies* 54 (3): 415–33.

Frye, Timothy. 2015. "What Do Voters in Ukraine Want?: A Survey Experiment on Candidate Ethnicity, Language, and Policy Orientation." *Problems of Post-Communism* 62 (5): 247–57.

Gawronski, Bertram, and Galen V. Bodenhausen. 2006. "Associative and Propositional Processes in Evaluation: An Integrative Review of Implicit and Explicit Attitude Change." *Psychological Bulletin* 132 (5): 692–731.

Gibler, Douglas M., Marc L. Hutchison, and Steven V. Miller. 2012. "Individual Identity Attachments and International Conflict: The Importance of Territorial Threat." *Comparative Political Studies* 45 (12): 1655–83.

Giuliano, Elise. 2015. "The Social Bases of Support for Self-Determination in East Ukraine." *Ethnopolitics* 14 (5): 513–22.

———. 2018. "Who Supported Separatism in Donbas? Ethnicity and Popular Opinion at the Start of the Ukraine Crisis." *Post-Soviet Affairs* 34 (2-3): 158–78.

Goble, Paul. 2016. "Russan National Identity and the Ukrainian Crisis." *Communist and Post-Communist Studies* 49 (1): 37–43. s.

Greenwald, Anthony G., Brian A. Nosek, and Mahzarin R. Banaji. 2003. "Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm." *Journal of Personality and Social Psychology* 85 (2): 197–216.

Greenwald, Anthony G., Brian A. Nosek, and Nene Sriram. 2006. "Consequential Validity of the Implicit Association Test: Comment on Blanton and Jaccard (2006)." *American Psychologist* 61: 56–61.

Greenwald, Anthony G., T. Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R. Banaji. 2009. "Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity." *Journal of Personality and Social Psychology* 97 (1): 17–41.

Hale, Henry E. 2011. "Formal Constitutions in Informal Politics: Institutions and Democratization in Post-Soviet Eurasia." *World Politics* 63 (04): 581–617.

———. 2018. "How Crimea Pays: Media, Rallying 'Round the Flag, and Authoritarian Support." *Comparative Politics* 50 (3): 369–91.

Herbst, Jeffrey. 1990. "War and the State in Africa." *International Security* 14 (4): 117.

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

Kam, Cindy D., and Jennifer M. Ramos. 2008. "Joining and Leaving the Rally: Understanding the Surge and Decline in Presidential Approval Following 9/11." *Public Opinion Quarterly* 72 (4): 619–50.

Kaufmann, Chaim. 1996. "Possible and Impossible Solutions to Ethnic Civil Wars." *International Security* 20 (4): 136–75.

Kobayashi, Tetsuro, and Azusa Katagiri. 2018. "The "Rally 'Round the Flag Effect" in Territorial Disputes: Experimental Evidence from Japan–China Relations." *Journal of East Asian Studies* 18 (3): 299–319.

Korostelina, Karina V. 2013. "Mapping National Identity Narratives in Ukraine." *Nationalities Papers* 41 (2): 293–315.

Korovkin, Vasily, and Alexey Makarin. 2019. "Conflict and Inter-Group Trade: Evidence from the 2014 Russia–Ukraine Crisis." Available at SSRN 3397276.

Kulyk, Volodymyr. 2011. "Language Identity, Linguistic Diversity and Political Cleavages: Evidence from Ukraine." *Nations and Nationalism* 17 (3): 627–48.

———. 2016. "National Identity in Ukraine: Impact of Euromaidan and the War." *Europe–Asia Studies* 68 (4): 588–608.

Kuran, Timur. 1991. "Now Out of Never: The Element of Surprise in the East European Revolution of 1989." *World Politics* 44 (1): 7–48.

———. 1998. "Ethnic Norms and Their Transformation through Reputational Cascades." *The Journal of Legal Studies* 27 (S2): 623–59.

Lai, Brian, and Dan Reiter. 2005. "Rally 'Round the Union Jack? Public Opinion and the Use of Force in the United Kingdom, 1948–2001." *International Studies Quarterly* 49 (2): 255–72.

Lambert, Alan J., Laura D. Scherer, John Paul Schott, Kristina R. Olson, Rick K. Andrews, Thomas C. O'Brien, and Alison R. Zisser. 2010. "Rally Effects, Threat, and Attitude Change: An Integrative Approach to Understanding the Role of Emotion." *Journal of Personality and Social Psychology* 98 (6): 886–903.

Lane, Kristin A., Mahzarin R. Banaji, Brian A. Nosek, and Anthony G. Greenwald. 2007. "Understanding and Using the Implicit Association Test: IV What We Know (So Far) about the Method." In *Implicit Measures of Attitudes*, edited by Bernd Wittenbrink and Norbert Schwarz, 59–102. New York: Guilford Press.

Lankina, Tomila, and Kohei Watanabe. 2017. "'Russian Spring' or 'Spring Betrayal?' The Media as a Mirror of Putin's Evolving Strategy in Ukraine." *Europe–Asia Studies* 69 (10): 1526–56.

Lyall, Jason, Graeme Blair, and Kosuke Imai. 2013. "Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan." *American Political Science Review* 107 (4): 679–705.

Malesky, Edmund J., Dimitar D. Gueorguiev, and Nathan M. Jensen. 2015. "Monopoly Money: Foreign Investment and Bribery in Vietnam, a Survey Experiment." *American Journal of Political Science* 59 (2): 419–39.

Martin, Terry. 1998. "The Origins of Soviet Ethnic Cleansing." *The Journal of Modern History* 70 (4): 813–61.

McNamee, Lachlan, and Anna Zhang. 2019. "Demographic Engineering and International Conflict: Evidence from China and the Former USSR." *International Organization* 73: 291–327.

Merolla, Jennifer L., and Elizabeth J. Zechmeister. 2013. "Evaluating Political Leaders in Times of Terror and Economic Threat: The Conditioning Influence of Politician Partisanship." *Journal of Politics* 75 (3): 599–612.

Mo, Cecilia Hyunjung. 2015. "The Consequences of Explicit and Implicit Gender Attitudes and Candidate Quality in the Calculations of Voters." *Political Behavior* 37 (2): 357–95.

Mueller, John E. 1973. *War, Presidents & Public Opinion*. New York: John Wiley & Sons.

Mylonas, Harris. 2012. *The Politics of Nation-Building: Making Co-Nationals, Refugees, and Minorities*. Cambridge: Cambridge University Press.

Nosek, Brian A., Mahzarin Banaji, and Anthony G. Greenwald. 2002. "Harvesting Implicit Group Attitudes and Beliefs from a Demonstration Web Site." *Group Dynamics: Theory, Research, and Practice* 6 (1): 101–15.

Ojeda, Christopher. 2016. "The Effect of 9/11 on the Heritability of Political Trust." *Political Psychology* 37 (1): 73–88.

Onuch, Olga, and Henry E. Hale. 2018. "Capturing Ethnicity: The Case of Ukraine." *Post-Soviet Affairs* 34 (2-3): 84–106.

Perrin, Andrew J., and Sondra J. Smolek. 2009. "Who Trusts? Race, Gender, and the September 11 Rally Effect among Young Adults." *Social Science Research* 38 (1): 134–45.

Petersen, Roger D. 2002. *Understanding Ethnic Violence: Fear, Hatred, and Resentment in Twentieth-Century Eastern Europe*. New York: Cambridge University Press.

Pérez, Efrén O. 2010. "Explicit Evidence on the Import of Implicit Attitudes: The IAT and Immigration Policy Judgments." *Political Behavior* 32 (4): 517–45.

Posen, Barry R. 1993. "The Security Dilemma and Ethnic Conflict." *Survival* 35 (1): 27–47.

Roccato, Michele, and Cristina Zogmaister. 2010. "Predicting the Vote through Implicit and Explicit Attitudes: A Field Research." *Political Psychology* 31 (2): 249–74.

Russett, Bruce. 1990. "Economic Decline, Electoral Pressure, and the Initiation of Interstate Conflict." In *Prisoners of War? Nation-States in the Modern Era*, edited by Charles S. Gochman and Alan Ned Sabrosky, 123–40. Lexington, MA: Lexington Books.

Ryan, Timothy J. 2017. "How Do Indifferent Voters Decide? The Political Importance of Implicit Attitudes." *American Journal of Political Science* 61 (4): 892–907.

Saideman, Stephen M., and R. William Ayres. 2008. *For Kin or Country: Xenophobia, Nationalism, and War*. New York: Columbia University Press.

Sasse, Gwendolyn, and Alice Lackner. 2018. "War and Identity: The Case of the Donbas in Ukraine." *Post-Soviet Affairs* 34 (2-3): 139–57.

Shevel, Oxana. 2002. "Nationality in Ukraine: Some Rules of Engagement:" *East European Politics and Societies* 16 (2): 386–413.

Sigelman, Lee, and Pamela Johnston Conover. 1981. "The Dynamics of Presidential Support during International Conflict Situations: The Iranian Hostage Crisis." *Political Behavior* 3 (4): 303–18.

Somer, Murat. 2001. "Cascades of Ethnic Polarization: Lessons from Yugoslavia." *The Annals of the American Academy of Political and Social Science* 573: 127–51.

Tajfel, Henri, and John Turner. 1979. "An Integrative Theory of Intergroup Conflict." In *The Social Psychology of Intergroup Relations*, edited by William G. Austin and Stephen Worchel, 33–47. Monterey, CA: Brooks/Cole PubCo.

Tellez, Juan Fernando. 2019. "Worlds Apart: Conflict Exposure and Preferences for Peace." *Journal of Conflict Resolution* 63 (4): 1053–76.

Theiler, Tobias. 2017. "The Microfoundations of Diversionary Conflict." *Security Studies* 27 (2): 318–43.

Tilly, Charles. 1990. *Coercion, Capital, and European States, AD 990–1990. Studies in Social Discontinuity*. Cambridge, MA: Basil Blackwell.

Toal, Gerard. 2017. *Near Abroad: Putin, the West and the Contest Over Ukraine and the Caucasus*. New York: Oxford University Press.

Toft, Monica Duffy. 2010. *The Geography of Ethnic Violence: Identity, Interests, and the Indivisibility of Territory*. Princeton, NJ: Princeton University Press.

Truex, Rory, and Daniel L. Tavana. 2019. "Implicit Attitudes toward an Authoritarian Regime." *Journal of Politics* 81 (3): 1014–27.

Wallace, Geoffrey P.R. 2015. *Life and Death in Captivity: The Abuse of Prisoners during War*. Ithaca, NY: Cornell University Press.

Wimmer, Andreas. 2013. *Waves of War: Nationalism, State Formation, and Ethnic Exclusion in the Modern World. Cambridge Studies in Comparative Politics*. New York: Cambridge University Press.

Wood, Elisabeth Jean. 2008. "The Social Processes of Civil War: The Wartime Transformation of Social Networks." *Annual Review of Political Science* 11 (1): 539–61.