When Groups Fall Apart: Measuring Transnational Polarization during the Arab Uprisings

and Robert Kubinec*and John Owen

Department of Politics

University of Virginia

October 6, 2019

Abstract

It is generally impossible to observe counterfactuals in which ideological polarization occurs both with and without international influence. To address this thorny problem, we employ a new statistical method, item response theory-vector auto-regression (IRT-VAR), that permits us to identify separately the transnational and domestic dynamics of polarization after the Arab Uprisings of 2011. We collected a dataset of Twitter accounts in Egypt and Tunisia during the critical year of 2013 when the Egyptian military staged a coup against the Islamist president Muhammad Morsi. We find that the coup increased retweets among ideological allies by as high as 50 percent each day following the coup and decreased cross-ideological retweets by as high as 25 percent. However, we also show that the transnational influence of external Islamist groups on Egyptian's Islamists served to dampen domestic polarization relative to the level that polarization might have reached if international social connections had not existed.[1]

*Corresponding author rmk7@nyu.edu

1. We thank Steven Livingston and meeting participants at the 2017 American Political Science Association Annual Conference, the 2018 Annual Conference on Politics and Computational Social Science, and New York University's Social Media and Political Participation Lab for valuable criticism. The authors thank the Amb. Henry J. and Mrs. Marion R. Taylor Chair at the University of Virginia for research funding. John Owen thanks the Liu Institute for Global Inquiry at the University of British Columbia for providing a writing venue. Robert Kubinec thanks the

Word count: 11,986

Introduction

International diffusion is a perennial and large question in comparative politics, straddling the intellectual boundary with international relations. When we observe similar policies, regimes, social movements, ideologies, or unrest across countries, are they caused by similar conditions in those countries, or have they spread internationally among the countries through some mechanism? The question implicates both our theories of politics and our empirical strategies. Galton's Problem captures the empirical issue: should observations of a phenomenon in several countries at once be treated as independent, or might autocorrelation be at work – for example, through transnational learning and imitation?

Skeptics of international diffusion remain, but the thesis that diffusion happens has strong momentum in the scholarly literature (Solingen 2012). Recent work has focused on empirical testing of theoretical claims about diffusion (Franzese Jr and Hays 2008). Our collective understanding of the phenomenon would advance greatly were we able to observe diffusion directly, but the actual international spreading of at least some categories of behavior is probably impossible. In this paper, we use social media to observe as directly as possible the diffusion of political contention across state boundaries using data from social media. Specifically, we investigate whether polarization between Islamists and secularists and democrats and authoritarians in Egypt and Tunisia was affected by events in Egypt in 2013, during the aftermath of the Arab Uprising.

Although early studies are skeptical of diffusion of the Arab Uprising itself (Aday et al. 2012), there are theoretical reasons to believe that group polarization should indeed diffuse across national boundaries in cases such as this, where the two countries, for all their differences, share a majority

University of Virginia's Quantitative Collaborative and New York University Abu Dhabi for computing resources and Jonathan Kropko for helpful feedback on this project. We thank Hana Nasser and Dana Moyer for invaluable research assistance.

religion and language, inhabit the same geographical region, and are open to transnational communication. Empirically, however, it is extremely difficult to disentangle the effects of those and other common conditions from actual diffusion of group conflict – that is, from people in Tunisia observing events in Egypt and moving further apart or closer apart ideologically as a result. In essence, we are missing a counterfactual in which we can observe domestic polarization both with and without transnational influence.

Our dependent variable is change in group polarization. Group polarization is situational and hence may be short-lived; it is different from the stable, long-term social polarization (e.g. Red versus Blue America) that many social scientists study (Moulaert and Sekia 2003). An endogenous process triggered by exogenous events, group polarization is a way to conceive of how identities and preferences change in response to stimuli such as a public demonstration or a coup d'état. A stylized version runs as follows: At time t, an unmodeled stimulus takes place (e.g., Mohamed Bouazizi's self-immolation in Sidi Bouzid, Tunisia); at $t + 1$, the news spreads; at $t + 2$, a portion of the population of Sidi Bouzid becomes angry and identifies more strongly against the local and national government, as reflected in anti-government speech and action; at $t + 3$, news of that speech and action spreads; at $t + 4$, a portion of Sidi Bouzid's population becomes angry at the protesters and identifies more strongly with the government, as reflected in pro-government speech and action; at $t + 5$, news of that speech and action spread within Tunisia and beyond; and so on. Group polarization is self-reinforcing. It may be slowed, halted, or reversed by various developments, including coercion (censorship, physical force) by governments.

Its situational quality, specifically its dependence on exogenous events, makes group polarization impossible to predict, but therefore especially helpful in understanding complex chains of events such as the Arab Uprising itself and the various political events that followed. The group-polarization approach presupposes, with a long tradition in social theory, that people have multiple group affiliations; that belonging to one group entails defining oneself over against one or more alternative groups; that for a given individual at a given time a particular group affiliation may be more or less salient; that individuals respond to signals of friendship and hostility from one another; and hence that populations sometimes polarize along one axis of identity, temporarily submerging other axes of identity. Of particular interest after the Arab Uprising are (1) polarization along one identity axis across a national population, and (2) transnational polarization, i.e., simultaneous polarization along an identity axis in two countries.

In this paper, we present analysis of Twitter data during 2013 in Egypt and Tunisia. We measure the degree of Islamist-secularist and democrat-authoritarian polarization on a given day by the proportion of retweets by ordinary citizens of ideological elites. (We identify the ideology of elites by the content of their tweets and known group membership.) A rising proportion of Islamist retweets among Islamists, for example, means that Islamism is becoming more salient for Islamists.

We choose this time period because Egypt went through a number of political events that seem prime candidates for triggers of group polarization. Most obvious is the July 3 coup d'état in which secularist military officers overthrew the elected government of Mohamed Morsi of the Muslim Brotherhood. In our models we specifically test for the effect of this coup (which we assume at least the timing of is exogenous) on endogenous polarization. We find that it has a statistically discernible long-term effect on the diffusion of group polarization across countries. Overall, we find that retweets from and to Islamists in Egypt and Tunisia increased by approximately 50% and 20% respectively following the coup, while retweets from and to secularists in Egypt and Tunisia increased by approximately 30% and 20%.

In addition to these empirical estimates, we also identify separately the *direct* effect of the coup on polarization in both countries from the *indirect* influence of transnational allies on polarization. This identification is achieved by virture of the model of latent polarization we explicate in this paper. As such, we believe it represents the first such estimates in the literature at this level of granularity, permitting us to learn about the dynamic nature of transnational polarization.

Group Polarization: An Informal Model

Polarization has been studied extensively by social scientists. Much of that work concerns social polarization, or segregation into groups that are stable over long periods of time (such as "Red

America" and “Blue America"). By group polarization, we mean a process of segregation – not an equilibrium – that is relatively short-term or situational yet may be politically consequential. Along with other scholars, we define polarization as a social construct, namely as progressive identity change that entails preference change. When two actors polarize, at time t, both actors may prefer a 50-50 allocation of goods; at $t + 1$, each may prefer a 60-40 allocation in its favor; at $t + 2$, a 75-25 allocation; and so on. At the limit of polarization, each side wishes the other destroyed.

Group polarization, then, is one way to formulate a progressive worsening of conflict; it does not cause conflict, in the sense of an independent variable causing a dependent one. Rather, polarization is conflict that is self-intensifying. Group polarization is endogenous, not in the sense that it is unrelated to pre-existing cleavages bur rather in the sense that, once triggered by exogenous events, it is self-exacerbating (Sambanis and Shayo 2013). It entails the altering of individuals' preferences and practices and creates new threats and opportunities for various actors, including actual and aspiring rulers (Owen and Inboden 2015, 55-61).

Stated informally, the basic group polarization model is simple. Assume a population of 100 persons, all belonging to one half or another of x pairs of opposing social groups (labor or capital, democrats or authoritarians, Islamists or secularists, urban or rural, etc., etc.). Fifty are pro-democracy, fifty pro-authoritarian. These groups do not correlate significantly to any other groups; e.g., democrats are as likely to be Islamist as secularist. The population thus has cross-cutting cleavages. At time t, the population begins in an unbiased equilibrium, such that, although individuals may identify more strongly with one group affiliation than with others, in the population as a whole, no identity axis predominates; hence social interaction does not skew the distribution of resources, including information, to any of the social groups (Dunning and Harrison 2010). Now suppose that at $t+1$ three democrats – one Islamist and two secularist, and two urban and one rural – publicly beat an authoritarian. Assuming a relatively free flow of information, that event can trigger the polarization of the population along a democratic-authoritarian ideological axis, such that democrats and authoritarians care less and less about class, being urban or rural, or mosque-state questions and more and more about ideology. The endogeneity of group polarization implies that we can distinguish a *direct* effect from the beating (people observe the beating, feel threatened, and begin to polarize) from an *indirect* effect (people observe their confréres polarizing, and do likewise). If not disrupted, polarization by definition culminates in inter-group violence.

Transnational group polarization takes place when polarization spans two or more countries at once. Citizenship in a state amounts to yet another group affiliation, albeit normally an especially politically salient one that carries the advantages of a state apparatus. States are set up to foster group identity and loyalty vis-à-vis foreigners. They may use physical segregation, closed or semi-closed national borders, national economic integration, propaganda, history, threats of war, coercion, and other means to induce a strong national identity among citizens and hence a strong sense that foreigners are an “other." Yet, interaction – communication, trade, investment, travel – across state borders is normal, particularly among most countries in the twenty-first century. States vary in their capacity to build and maintain a national identity and to have that identity perpetually trump all other group affiliations, including transnational ones, across all conditions. Thus transnational group affiliations – ethnic, religious, ideological, class, sexual – are part of life for most people in most countries. Insofar as communication across state boundaries is uncensored by states, transnational group affiliations can yield transnational group polarization, as a direct effect (from observing the stimulus) produces an indirect effect (from observing others polarize). The informal model above may then incorporate democrats and authoritarians in a second state (and a third, a fourth, and so on).

Justifying Assumptions

Social identity theory links the formation of groups and their degree of competition by means of the concept of polarization. Microfoundations for such a model are found in philosophy and social theory. Assume that persons are not atomized individuals whose fundamental goal is to maximize their own exogenously derived utility and who value the gains and losses of others only insofar as those are instrumental to such maximization. Assume instead the persons depicted by traditions in sociology (Simmel 1955; Coser 1956): each individual is fundamentally a member of multiple social groups, and he identifies his interests to some extent with those of the groups to which he belongs and against opposing groups.

The logical foundations of this communal psychology is seen in the formula articulated by Spinoza and, later, Hegel: omnis determinatio est negatio, or “all determination is negation" (Melamed 2013). A thing must necessarily have properties, such as “short" or “cold." But properties only exist in contrast to other properties ("tall" or "hot") (Taylor 1975, 232-239). Having a property is equivalent to belonging to the set of things that have that property (Quine 1989, 22-26). Being female is equivalent to being a member of the set of persons that are female. Identity thus is social: who I am entails my group memberships.

Experimental evidence suggests that people, or at least some people, tend to think, feel, and act according to this in-group versus out-group logic. People tend to be self-interested, but their notion of “self" may expand to include persons in their social group whose existence requires contrast with some opposing or “out-group" (Mercer 1995). Indeed, these two identifications are mutually constitutive. As Simmel put the matter,

> It appears to be necessary for us human beings, whose whole psychical nature is built upon our sensitiveness of difference, that a feeling of separateness should always exist alongside of the feeling of unity to make this latter perceptible and tangible (Simmel 1898, 45-46).

Overlapping Social Groups and Different Saliencies

That individuals belong to multiple social groups, each with a corresponding anti-group, introduces a complication. For Simmel, an individual’s identity consists of the unique overlap of the groups to which she belongs (Simmel 1955, 139-41). Yet, at any given moment a given individual will identify more strongly with some of his groups than with others. Sometimes large groups of individuals may place greater weight on one particular group affiliation simultaneously, such that populations polarize along a particular axis of identity. Social-psychological literature posits at least two attributes of groups that lend them high salience. One is prestige or high status: experiments show that members of high-status groups are significantly more biased toward fellow members and against nonmembers than are members of low-status groups. Experiments also show that a second attribute is threat (physical, economic, status, etc.) – particularly among persons already highly committed to the group (Ellemers et al. 2002). A new threat – such as an attack on a group member by members of the opposing group – tends to arouse in such persons fears that they may be next, and so they tend to increase their biases toward that particular group affiliation. They identify more with it and against the threatening group. This experimental result is anticipated by Simmel:

> It is a fact of the greatest social significance, one of the few which are true almost without exception of group formations of every sort, that common antagonism against a third party under all circumstances tends to consolidate the combining group, and with much greater certainty than community in friendly relationships toward a third party (Simmel 1898, 45-46).

If these attributes of prestige and threat are associated with high salience, it should be the case that a rise in a group's status or jeopardy can render it more salient for its members. A rise in status may be triggered in politics by a victory in an election or a civil war, or an unexpectedly large public rally. A rise in threat may be brought on by physical violence, verbal abuse, or evidence (true or false) of discrimination or persecution against the group.

Social-psychological literature notes that people vary by level of commitment to various groups. In equilibrium some city-dwelling Islamists identify more as urban and less as Islamist; others identify more as Islamist as less as urban. This kind of heterogeneity could in principle stifle polarization, because low-commitment group members could try to exit or hide from the group rather than take the risks that come with strongly identifying with it. Against that possibility, Tilly writes that, following a triggering event, highly committed group members mediate and broker polarization by spreading information about the threat or increased status and about ongoing polarization. Such brokers may propagandize by exaggerating and inventing symbolic events. Public discourse turns to what is to be done; those who hold extreme views tend to have more influence in such times and moderates either are quiet or move toward the extreme (Tilly 2005, 143-44). Smith (2012) models allocation decisions in a game comprising two social groups, each comprising two types of actors: “behavioral" actors who are biased to favor their own group members, and “rational" actors who are unbiased. The model shows that “rational" across will come to act like “behavioral" ones and favor allocation of goods to their own group.

It stands to reason that the low-commitment actors posited by Ellemers, Kortekaas, and Ouwerkerk (1999) would behave like Smith’s “rational" actors. In sum, an exogenous event that either raises the prestige of social group A or threatens group A may cause people who belong to multiple overlapping groups A through Z to identify more strongly with A and against $\neg A$ and less with B and against $\neg B$, etc. Increases in status and in threats may be simultaneous: an increase in A’s status may simultaneously threaten members of $\neg A$ and thus cause them to identify more as $\neg A$s and against group A. Large public anti-government demonstrations, as take place during a typical political spring, can both raise the status of being anti-government and simultaneously threaten those who identify with the government. And again, polarization tends to feed on itself: as members of A observe members of $\neg A$ identifying more as $\neg A$s, members of A will identify still more strongly with A; and so on, in what we call an indirect effect.

For that reason, we are interested in establishing that, in general, (1) relatively exogenous events like the coup against Mohamed Morsi in July of 2013 will affect the direction of group polarization, and (3) these exogenous events also induce endogenous (de-)polarization as groups’ polarization increases (decreases) in response to their ideological allies’ polarization. We are not trying to ascertain conditions under which polarization does not diffuse – only whether it sometimes does so. Thus we select on the explanatory variable of an exogenous stimulus of polarization.

Our two countries are Egypt and Tunisia. It is important to have theoretically informed reasons for assuming that two states are candidates for international diffusion (Zhukov and Stewart 2013). We choose Egypt and Tunisia because they are part of an ex ante identifiable cultural, ethnic, and geographical region and the two were open to some of the same media, including satellite television and Internet platforms. Evidence from interviews suggests that our general expectation of polarization following catalytic events is plausible. Concerning transnational polarization, Marks reports that the mainstream Islamist party in Tunisia, Ennahda, traditionally identified little with Egypt's Muslim Brotherhood, looking instead to Turkey's Justice and Development Party (AK Parti) as a model; but that the July 3, 2013, coup against Mohamed Morsi in Egypt – the Muslim Brotherhood leader who had been elected president – generated new sympathy in Ennahda for their Islamist counterparts in Egypt and made them more suspicious of secularist Tunisians (**4**; Marks 2015).

Hypotheses

We propose to test the following hypotheses based on this theory:

- H1 After the coup against Mohamed Morsi, in each country the difference in latent ideological positions between competing groups will diverge in direct reaction to the coup (direct effect).
- H2 After the coup against Morsi, competing groups will show additional influence from the reactions of like-minded allies in other countries (indirect effect).

We operationalize these hypotheses by estimating the latent space within which people identify themselves with respect to these polarizing cleavages. To accomplish this, we collected data on ideological actors in Egypt and Tunisia from Twitter, as we described in the next section. To test these hypotheses, we need to be able to trace out the endogenous processes of polarization; we must address what Solingen (2012, 638) calls "the challenge of establishing when outcomes can be considered outcomes."

The event that we can identify a priori as potentially polarizing is the coup that overthrew Egypt's Islamist President Mohamed Morsi on July 13, 2013, as it seemed likely that the average secularist, the average Islamist, the average democrat and the average authoritarian would react differently. Other polarizing events occurred and our model should be able to identify them, but we focus our attention first on the coup because we believe based on prior information that it should be a directly polarizing phenomenon as it removed Islamists from power and ended democratization in Egypt. However, to date no one has been able to say for sure how much and to what extent Morsi's coup affected group polarization because group polarization in these countries was and is an ongoing endogenous process, making it very difficult to disentangle cause and effect.

Data Collection

To obtain our data, we started with a universe of tweets from the early stage of the Arab Uprising, December 2010 to April 1st, 2011 that all matched the search keywords "Cairo", "Alexandria" and "Tunis" in the user self-reported location field in Twitter.[2] While this dataset comprised 11 million tweets, it nonetheless did not capture all the important or influential users because it is quite common for Twitter users to either not report their location or to list a location that is not geographic in nature. For these reasons, to identify influential users who were not in this sample, we parsed the tweets in order to identify those users that had received the largest number of retweets and mentions during that time period. In this way, even if an influential user was not a part of the original sample, we were able to locate most of the popular Twitter users in the country by analyzing the content of the tweet database.

We curated the resulting list of elite users, both by removing accounts that were later abandoned and by adding in accounts that were created later in time. In general, Twitter in 2011 was dominated by well-educated secular elites with a flair for democracy, while Islamists and pro-regime figures were later to adopt Twitter as a medium. For example, in Egypt we added the account for Mohamed Morsi, the Islamist president affiliated with the Muslim Brotherhood, and in Tunisia we added Rached Ghannouchi, long considered to be the guiding force of the Islamist Nahda movement in Tunisia. The full list of users selected for analysis is available in the appendix.

2. The particular reason we started with this dataset is because we had access to these tweets from a prior research project. These tweets were purchased through the Gnip corporation (now owned by Twitter).

The final sample amounted to 148 Twitter users, 56 from Tunisia and 92 from Egypt. The larger number from Egypt reflects the much larger Twitter-sphere in the country and hence the need to obtain a broader sample of users.

With this finished list of 1 elite users, we then collected their full tweet histories from March 31st, 2013 to December 31st, 2013. This 1.7 million list of tweets was then filtered down to 1.2 million tweets that constitute each user's own composed tweets by removing all of each user's retweets. Using Twitter's open API, the retweets of all of these 1.2 million original tweets was then downloaded as a list of user IDs for each user per day for a total of 275 days. The use of the open API represents a limitation in the data collection because only 100 re-tweets of a given retweet may be downloaded; however, this limit was rarely reached in practice because very few of the tweets in question had more than 100 retweets. Even ex-President Morsi, who had more than a million Twitter followers, averaged only a few hundred re-tweets per tweet during 2013. Nonetheless, these data must still be understood as a sample of the full number of retweets, especially for users with a very large popular following.

In the next stage we coded all of our users along our two central latent scales of interest: Islamism/secularism and democracy/authoritarianism. We had two graduate assistants code the user list along two axes: Islamist-Secularist and Pro/Anti Democracy. The graduate assistants used polarizing search words on the elites' Twitter feeds, such as common hash tags in the Arab Uprising (Steinert-Threlkeld 2017), and also looked at published writings or newspaper articles about these users. If a user did not have any political content that expressed their own opinions they were removed from the sample (primarily media accounts).

Ultimately we found elite users at almost all possible combinations of religion and democracy perspectives. While Islamist and secularist were relatively easy to diagnose given the heightened salience of Islamist groups in politics during this time period, diagnosing an elite's views on democracy proved to be a significantly greater challenge. The reason for this difficulty is because of the strong normative bias against expressing pro-authoritarian discourse during this time period, although some users were willing to tweet their skepticism of democracy and the Arab Spring.

	Democracy Coding	Religion Coding
Percent Agreement	38.9% (83)	61% (130)
Percent Disagreement	72% (77)	28% (30)

Note: Rows sum to one.

Table 1: Coding Agreement for Anti/Pro Democratic and Islamist-Secularist Twitter Users

For others, we were able to determine their anti-democratic views based on their membership in pro-authoritarian parties in Egypt and Tunisia. However, reliability for coding democracy was too low to rely solely on prior information to form groups.

Table 1 illustrates this tendency for the coders to agree much more on the Islamist-secularist coding than on the pro/anti democracy coding. While the coders disagreed on 77 of the users, or roughly half, for the pro/anti democracy cleavage, they only disagreed on 30 of the Islamist/secularist cleavage codings. Furthermore, almost all of the remaining disagreements for Islamism and secularism were easy to resolve because they resulted in a coder failing to see a tweet or piece of evidence that another coder found. For those users who were the most difficult to characterize and had relatively little in the way of identifiable sectarian content, we defaulted to secularism given that the majority of the users tended to be secular in terms of appearance, education and language. In addition to these binary classifications, we also had the graduate coders record their confidence in their assessment on a scale of 0 to 100. We further reviewed cases that had an uncertainty of less than 50 percent even if both coders agreed in their original assessment. In general, we found that these users did not tweet as much on political topics and their ideology is relatively unknown. We excluded such users for the analyses we report here as their lack of any political content makes them uninteresting to this analysis.

Because of the difficulty in using prior information to code all users for pro/anti democratic membership, we instead implemented a static 2-dimensional item response theory model (Clinton, Jackman, and Rivers 2004) where we coded a handful of users whose pro/anti democratic sentiment was well-known while including our pre-existing secular/Islamist coding as the first dimension. The data we used for this model are the same data we employed for our full analysis later in this paper except that retweet counts were aggregated over time. We allowed the ideal points of all the users in the model to float along the democracy dimension, and then we classified users as belonging to the pro or anti-democracy group depending on whether their median posterior estimate was greater or less than the overall median for the latent scale. As such, this measurement strategy exploits our quality prior knowledge while avoiding having to make difficult-to-justify decisions about the vast majority of users.[3]

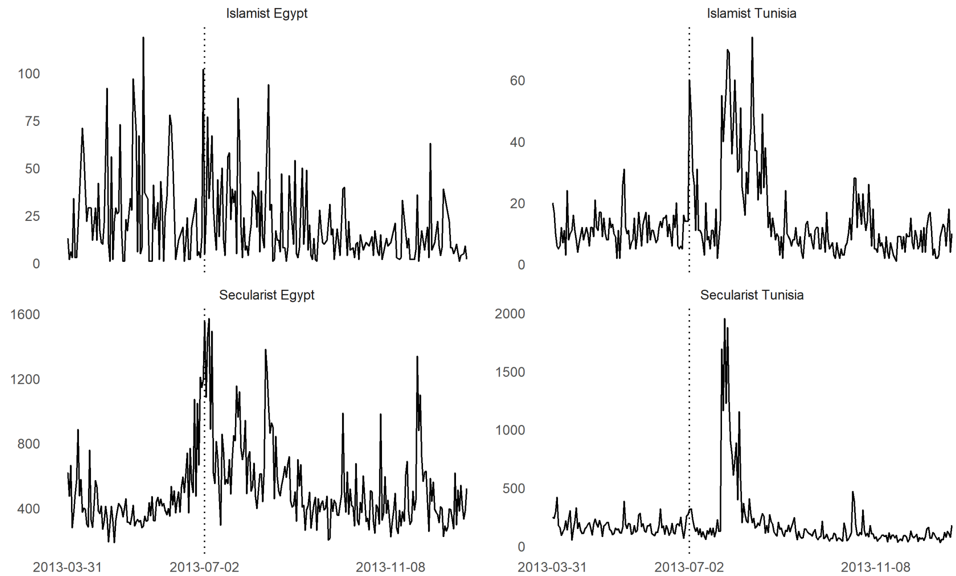
The final database comprises a set of elites j and citizens i in which the outcome is the number of times that i retweeted j for each 24-period in the sample.[4] We took these retweets and we calculated how many times a specific user/citizen retweeted each elite within the 24-hour period. We then appended these data together for all the elites in our sample. We removed all citizens who did not re-tweet at least five different elites during the entire time period, resulting in a final dataset of 119,470 citizen-elite interactions with a total of 6,134 unique citizen Twitter users, or an average of 19 retweets per user of elites during the sample period. We then expanded this dataset by including all absent interactions in the dataset; i.e., for every 24-period, we include zeroes for all elites which a citizen did not retweet. This dramatically expands the volume of data to 11,807,950 observed and possible interactions. Including zeroes is important so that we do not assume that each citizen had an equal chance of retweeting all the elite Twitter accounts in the sample, as we explain in our modeling strategy.

Figure 1 shows the number of retweets of elite users aggregated by sectarian affiliation and each day in the sample. As can be seen, Twitter account activity varies tremendously over time, with prominent spikes around July when the Egyptian military overthrew the Muslim Brotherhood

3. One limitation arising from this model is that it did not identify any Egyptian Islamists who were coded as anti-democratic. This lacuna is a feature of the data rather than the model; the Islamists prior to the coup were democratically elected and thus were on balance very supportive of the system of government.

4. As has become more well-known, automated Twitter accounts known as bots are able to inflate follower counts and amplify users' messages. However, because our inference is based on the curated list of elite users–none of whom are bots–the issue of bots is of minimal concern in this analysis as the measurement model is robust to purely mechanical retweeting. In addition, at this earlier stage in Twitter's development, bots were much less likely to be used as a tool of statecraft as is now more often the case.

Figure 1: Count of Retweets by Ideological Group Over Time



president Mohamed Morsi.

Modeling Ideological Polarization Over Time

We are not the first to collect and analyze Twitter data from the Middle East. Rather, our primary contribution in the empirical study of polarization is in our modeling strategy, which is broadly applicable to the study of these difficult-to-define processes. This new model, which we term item-response theory-vector autoregression (IRT-VAR), provides single summary measures of the influence of competing ideological groups that robustly captures the substantial uncertainty in how to best measure group affiliation and ideological polarization.

Understanding measurement uncertainty is very important for Twitter analysis, and has been approached in a variety of ways by existing studies. Generally speaking, scholars have applied some kind of model or aggregation algorithm to Twitter data before running statistical models, such as sentiment coding (Jamal et al. 2015; Siegel et al. 2018), network statistics (Freelon, Lynch, and Aday 2015) or counting the use of particular keywords or hashtags (Weber, Garimella, and Batayneh 2013). However, these methods have the limitation in that it is hard to say whether the particular aggregation rules show strong validity to the concepts that the study wants to analyze (Gerring 2012). The main problem is that scholars in political science are often not interested in the linguistic usage of terms on Twitter per se, but rather in *what these terms signify about unobserved social and political constructs*.

For this reason, some of the most successful work incorporating social media to date has been precisely that in which the observable Twitter characteristics are the study domain, such as to the diffusion of known hashtags or keywords (Steinert-Threlkeld 2017). An alternative approach is to create very large corpora coded by human beings in which the size of the corpora will permit more nuanced inference on the latent concepts without needing to parameterize the text/concept relationship (Mitts 2019; Siegel et al. 2018). The study of ideological polarization requires an assumption that the aggregation used accurately reflects the underlying social process that the analyst wishes to identify. When it comes to identifying latent social cleavages, the observed Twitter data are rarely of interest, but rather whether and to what extent the observed data provide information on latent cleavages. While we know that this information exists, it can be frustratingly hard to extract.

For this reason, even as Twitter data open up new opportunities for studying group formation processes in near real-time, they also present imposing hurdles because the medium is not designed for easy interpretation or classification. The short character limit on tweets and the way in which tweet replies are structured invite users to write tweets using sophisticated (or at the very least obtuse) irony and sarcasm, hence the need for very large corpora for machine-learning style classification.

Furthermore, discerning the meaning of tweets is doubly difficult in Arabic-language Twitter as existing computational linguistic models are usually limited to Western languages, and regional variants of Arabic, in addition to its non-standard Roman transliterations, make it hard to apply sentiment analysis. As a result, decisions over how to aggregate and measure tweets can have a strong influence on the results, and also make it difficult to compare across studies given the disparity in measurement strategies.

To address these concerns, we focus on a single type of Twitter-based behavior–retweets–and design a model that combines measurement with direct statistical analysis of group polarization. By so doing, our model only has to make two basic assumptions: 1) retweets are a signal of underlying ideological agreement and 2) our codings of elite Twitter users as Islamist or secularist are accurate. The first assumption has been shown to be valid through analyses of the political content of retweet networks (Conover et al. 2011). Every time a Twitter user clicks on the retweet icon by a tweet, they immediately broadcast it to their followers, signaling that the user believes this tweet is a message worth amplifying to their network.[5] The second assumption of coding elite Twitter accounts, while still an assumption, is relatively transparent and easy to replicate, and we include the entire set of coding decisions in the appendix as a reference. All other analysis of ideology that we present in this paper flows from the model and incorporates our full uncertainty in accurately identifying the underlying social process of polarization relative to the irrelevant noise

5. Even though some users will say that their retweets do not signal agreement, in general it is rare to observe personal accounts retweeting ideological enemies, as it is counter to the very idea of a personal account to amplify views that are reprehensible to one's own.

in the content of users' tweets. Our use of joint measurement/inference models permit us to test an important null hypothesis concerning our measurement strategy–whether our concepts in fact capture an important dimension of variation–while simultaneously permitting us to learn about dynamics in the world.

To create this new model, we develop a statistical mélange based on ideas from two distinct literatures in statistics: time-series econometrics, in particular vector auto-regression, and the item-response theory literature on estimating latent concepts. Item-response theory has been used in political science for estimating the latent positions of actors based on roll-call voting datasets (Clinton, Jackman, and Rivers 2004), and more recently, through social media and campaign finance contributions (Bonica 2014; Barberá 2015). Item-response theory (IRT) has also been applied to difficult measurement problems, in particular the construction of democratization indices from a variety of coding sources (Coppedge et al. 2017; Treier and Jackman 2008). Item response models estimate latent traits for individual cases that can be divided into two distinct groups, such as raters and countries, lawmakers and bills, or in our sample, citizen and elite Twitter users.

The main advantage of these models is that they are designed to provide latent estimates of fundamentally unobservable quantities, such as political ideology, without assuming that the measures are valid a priori. For that reason, there is a strong connection between IRT and factor analysis (Takane and Leeuw 1986), and similar concerns over the interpretation of the latent scores are justified. The latent dimension estimated by an IRT model will be the lowest-variance explanation of the observed data, but further prior knowledge and post-estimation validation is necessary to confirm that the results correspond with the concept of interest. We are confident in our application of this technique to the data because of our prior coding of users into similar ideological groups. For that reason, the resulting estimates will reflect these latent cleavages instead of arbitrary social behaviors or groupings. In other words, we anchor the model in our prior knowledge about elite users, and rely on the retweet patterns of ordinary users to see how these ideological networks are changing over time. As such, we define the latent scale in our model as inter-group salience rather than ideology per se because we are examining group polarization (a relatively short-lived phenomenon) instead of changes in the content of the ideologies.

The existing application of time series modeling to IRT is limited despite the fact that political science has many time-varying variables with considerable measurement error. This paper builds on the approaches of Martin and Quinn (2002) and Kropko (2013), who use random-walk priors on latent-traits to incorporate change in ideal points over time. In time-series lingo, these latent traits become integrated variables that exhibit an infinite memory process: any shifts in the ideal points are remembered in the time series over time (Box-Steffensmeier et al. 2014, Ch. 5). While this type of autocorrelation is appropriate for long-term time series, our focus on short-run dynamics suggests we use a model that is based around stationary ideological groups, i.e., that groups receive temporal shocks causing heightened salience of ideological divides for a short period of time and then return to an average level of social affinity (Ch. 2).

Finally, in addition to examining change over time, our research explores transnational ideological polarization. We want to know whether an increase in polarization in Egypt causes changes in polarization in Tunisia. For that reason, we need to look at multivariate time-series techniques, i.e., vector autogression (VAR). Vector autoregression involves the estimation of lags of different time series in the same equation (Sims 1980), and has been used in political science to study time series that interact with each other for several decades (Freeman, Williams, and Lin 1989). Our main innovation in this paper is to jointly fit a VAR and an IRT model so that the VAR can fully incorporate the measurement error present in the Twitter data.

The primary purpose in employing a VAR as the inferential method is that it enables us to track the endogenous feedback effects of the ideal points of transnational groups on each other over time. To set up this model, we start with two time series that represent the latent ideal points of different religious or political groups: $y_{c,g,t}$ and $y'_{c',g,t}$. These series are observed at discrete time units t and each belong to the same religious affiliation $g_1 \in \{Secularist, Islamist\}$ or political affiliation $g_2 \in \{Democratic, Authoritarian\}$ but different countries $c \in \{Tunisia, Egypt\}$. While we have two dimensions in our model (religion and democracy), we suppress the dimensional subscript and present the model in terms of the first dimension for simplicity (i.e., only one group g_1). In a VAR framework, we can use the following equation to signify two series (y and y') of the same group g but separate countries (c and c') each other through their prior period lags. The parameters

$\beta_{c,g,I}$ and $\beta_{c',g,E}$ control the relative influence of prior period lags for internal I influence (prior lag of same series) and external E influence (prior lag of group in other country):

$$y_{c,g,t} = \gamma_{c,g} + \beta_{c,g,I} y_{c,g,t-1} + \beta_{c',g,E} y'_{c',g,t-1} + \beta_{c,g,x} X_t + \epsilon_{c,g,t} \tag{1}$$

$$y'_{c',g,t} = \gamma_{c',g} + \beta_{c',g,I} y'_{c',g,t-1} + \beta_{c,g,E} y_{c,g,t-1} + \beta_{c',g,x} X_t + \epsilon_{c',g,t} \tag{2}$$

To make the model stochastic, we include $\epsilon_{c,g,t}$ and $\epsilon_{c',g,t}$ as white noise (stationary) errors so long as $\beta_{c,g,I}$ and $\beta_{c,g,E}$ meet the VAR stability conditions (Zivot and Wang 2006, 386-387).[6] So long as these parameters meet the stability conditions, the latent ideal points will over time return to their long-run equilibrium value $\gamma_{c,g}$ (i.e., the intercept). Substantively, these parameters provide estimates of how quickly a time series will return to its long-term mean given an exogenous shock ($\beta_{c,g,I}$) and the strength of influence of the other time series ($\beta_{c,g,E}$).[7]

We also include an additional parameter in each series, $\beta_{g,c,x}$. This parameter does not vary over time and rather represents the effect of the exogenous event X_t, which equals 1 after the coup against President Morsi in Egypt and 0 before the coup. As such, we can use $\beta_{g,c,x}$ as a direct measure of the long-term polarizing effect of the coup on each of the series. A null hypothesis of no effect of the coup would be the case in which $\beta_{g,c,x} = 0$.

Given that we have two groups and two countries, we have two sets each of ideal points series $y_{c,g,t}$ and $y'_{c',g,t}$ with two dimensions, which comprises a seven time-series system (Tunisian and Egyptian secularists, Islamists, democrats and authoritarians).[8] While we could pair each series with every other series, we instead chose to only pair each ideological group with their ideological allies in the other country. We impose this restriction because we aim to identify the effect

6. Loosely speaking, a VAR is stationary if the eigenvalues of the coefficients of the lags in the VAR have a modulus less than one. In essence, if these coefficients are too large in absolute terms relative to each other, the VAR will move away from its equilibrium level in either a random or explosive direction.

7. We only include one lag of each series in our VAR equation because the time series are themselves latent variables, and as such the normal lag selection procedures for VARs do not apply. We would prefer as well to keep the model as parsimonious as possible.

8. As mentioned elsewhere, we do not have Egyptian Islamists who are also authoritarian.

of transnational polarization, and also because the within-country groups are separately related through the IRT model that we explicate in the appendix. In other words, the within-country groups are implicitly scaled relative to each other, we just do not add parameters that would measure the influence of prior period lags within countries across groups. Again, this is primarily for reasons of model parsimony.

Given this framework, we are able to track endogenous changes in the latent ideal point time series via the coefficients $\beta_{c,g,I}$ and $\beta_{c,g,E}$ that provide evidence of how reactive groups are to their own prior ideological position and to the position of other groups. A small value of $\beta_{c,g,E}$ will imply that a religious (political) group is relatively insensitive to the actions of their foreign co-religionists, and a larger value that the religious (political) group is very sensitive to what happens in other countries. On the other hand, a small value of $\beta_{c,g,I}$ implies that a religious (political) group is relatively insensitive to its own prior position in the latent space. As such, a low value of this parameter also signifies the stability of the series. Supposing that the religious group experienced some kind of shock that heightened the salience of religious divides, a lower $\beta_{c,g,I}$ will imply that the group return quickly to the long-term mean of the series (the intercept $\gamma_{c,g}$). By contrast, a higher value of $\beta_{c,g,I}$ implies that the religious group is unstable and is unlikely to return quickly to a long-term mean.

However, we need to obtain the ideal point series themselves (y_t and y'_t) in order to calculate all of these effects. As we mentioned earlier, we cannot directly measure ideological agreement on Twitter. We could try to create relevant time series by aggregating our Twitter data around the content of tweets using particular key words or hashtags, such as the Arabic phrases for sectarian terms. However, as we mentioned earlier, that would make our analysis heavily dependent on the particular set of keywords we chose to employ. For that reason, we turn to item-response theory (IRT).

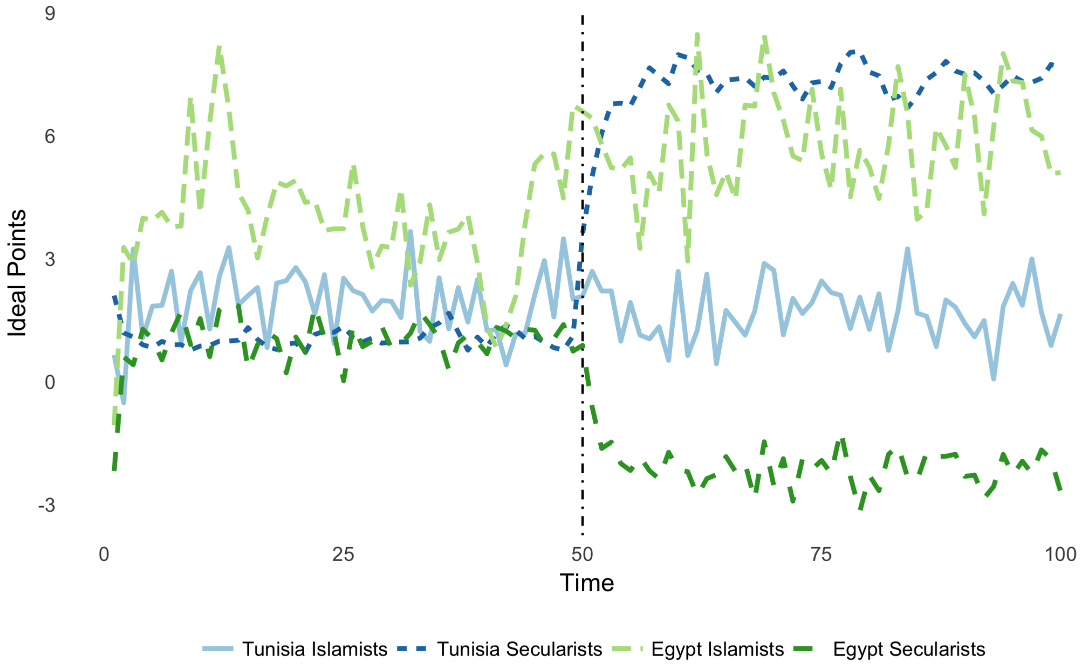
We provide a full explication of how we combine IRT and the VAR in Appendix B as well as explain how we can estimate this model using Bayesian inference. In essence, we use IRT to create the ideal points time series which are then jointly estimated with the VAR. IRT converts the raw Twitter data into four latent variables that represent the group-level ideal points of each of the religious-country groups. For these reasons, hereafter we refer to the time series $y_{c,g,t}$ and $y'_{c',g,t}$ as $\alpha_{c,g,t}$ and $\alpha'_{c',g,t}$ to reflect the fact that we are estimating latent, rather than observed, time series.

Because our IRT model employs the ideal point formulation (Clinton, Jackman, and Rivers 2004), the resulting latent variables can be interpreted as the relative salience of the groups in a latent social space in which users are making decisions about whom to retweet. Distance in this space then represents the closeness of these groups to each other in terms of ideologically-inclined retweet patterns. For example, as group polarization increases within countries, users will only retweet those elites who are close to their ideological ideal point, and we would expect the group-level ideal points to move farther apart in the latent space. Empirically that polarization would be expressed as more strongly polarized retweet patterns, i.e., whether or not Islamist elites receive more retweets from Islamist citizens, and vice versa for other groups. The IRT model is thus able to reduce our high-dimensional and rather noisy data down to a single estimate of relative group polarization with a credible interval to indicate our uncertainty in its true location. In other words, the group polarization estimate represents how users might think about and classify different people as belonging to these ideological groups, with periods of heightened polarization corresponding to more cleanly defined groups.

To test the model, we simulated data by generating group-level ideal points corresponding to the VAR equation above. The simulated "true" latent time series, each of which is paired with another time series representing an ideological group in different country, is shown in Figure 2. As can be seen, the two sets of time series each follow a pattern over time that is roughly stationary.

The vertical line in Figure 2 shows the time period where an exogenous event X_t occurs and $\beta_{g,c,x}$ has a non-zero value, which will shift the long-run equilibrium of each time series depending on the value of $\beta_{g,c,x}$. In addition to the summary estimate of $\beta_{g,c,x}$, we can also use the values of $\beta_{c,g,I}$ and $\beta_{c,g,E}$ to calculate impulse-response functions (IRFs) for a shock to the elite group's ideal points coming from the group's own time series, or the indirect effect coming from a shock to a different group's time series. We can express this mathematically as the derivative of an exogenous

Figure 2: Simulated Data for the Vector Auto-regressive Component of IRT-VAR



shock $\beta_{c,g,x}$ with respect to the value of the ideal point $\alpha_{c,g,t}$ at time points after the shock from $t+n, n \in \{1, 2, ...10\}$:

$$\frac{\partial \alpha_{c,g,t+n}}{\partial \beta_{c,g,t+n}} \tag{3}$$

This IRF essentially measures the decaying (if the time series is stable) average effect of a shock to the latent ideal points over time, and provides a straightforward measure of the effect of our explanatory variable on the outcome over time. We can also use this same framework to calculate another important IRF of interest, which is how the effect of the coup on the transnational group will affect the domestic group's ideal points:

$$\frac{\partial \alpha_{c,g,t+n}}{\partial \beta_{c',g,x}} \tag{4}$$

Each of these effects incorporates the endogenous nature of these processes, allowing for heightened polarization in prior time periods to influence the present, ad infinitum. To summarize the model, then, we can match these estimates to our hypotheses to provide very specific null hypothesis tests of our arguments that, as we have noted, incorporate into the tests our measurement uncertainty in using Twitter data. We show how this model relates to our hypotheses in Table 2.

Table 2: Hypotheses and Associated Tests from our Model

Hypothesis	Definition	Expected Outcome
H1	After the coup against Mohamed Morsi, the difference in latent ideological positions between Islamists/secularists and democrats/authoritarians will diverge in direct reaction to the coup (direct effect).	$\frac{\partial \alpha_{c,g,t+n}}{\partial \beta_{c,g,x}} \neq 0$ for all $c \in C$ and $g \in G$
H2	After the coup against Mohamed Morsi, the difference in latent ideological positions between Islamists/secularists and democrats/authoritarians will diverge from each group's reaction to their ideological allies' shift in latent ideological position (indirect effect).	$\frac{\partial \alpha_{c,g,t+n}}{\partial \beta_{c',g,x}} \neq 0$ for all $c \in C$ and $g \in G$

In the appendix we also discuss how we employ the method of Kubinec (2017) to adjust the ideal point estimates for self-selection by adding a 2-stage zero-inflation model. This adjustment is very important for accurate estimation because Twitter users tend to self-select which tweets they are likely to see on a given day because only those elites they have previously chosen to follow on Twitter will show up in their Twitter feed. As a result, we need to distinguish between "true" zeroes, in which a Twitter user actually saw a tweet and chose not to retweet, versus "false" zeroes, where a user never actually saw the elite's tweet and thus was never truly able to choose to retweet it.

While we reserve the discussion of this part of the model for the appendix, it is an important component of the model due to the serious sparsity in the data. If all we incorporated were raw retweet counts, we would dramatically over-state polarization by assuming that every user had scanned through all elites' Twitter feeds each day of the sample and decided whom to retweet on that day. Instead, we make the much weaker assumption that the probability of a user seeing an elite's tweet is likely first influenced by that user's position in the ideal point space, whether due to the user's prior decisions concerning which accounts to follow (Barberá 2015) or due to Twitter's recommendation algorithm suggesting like-minded accounts.

Model Results

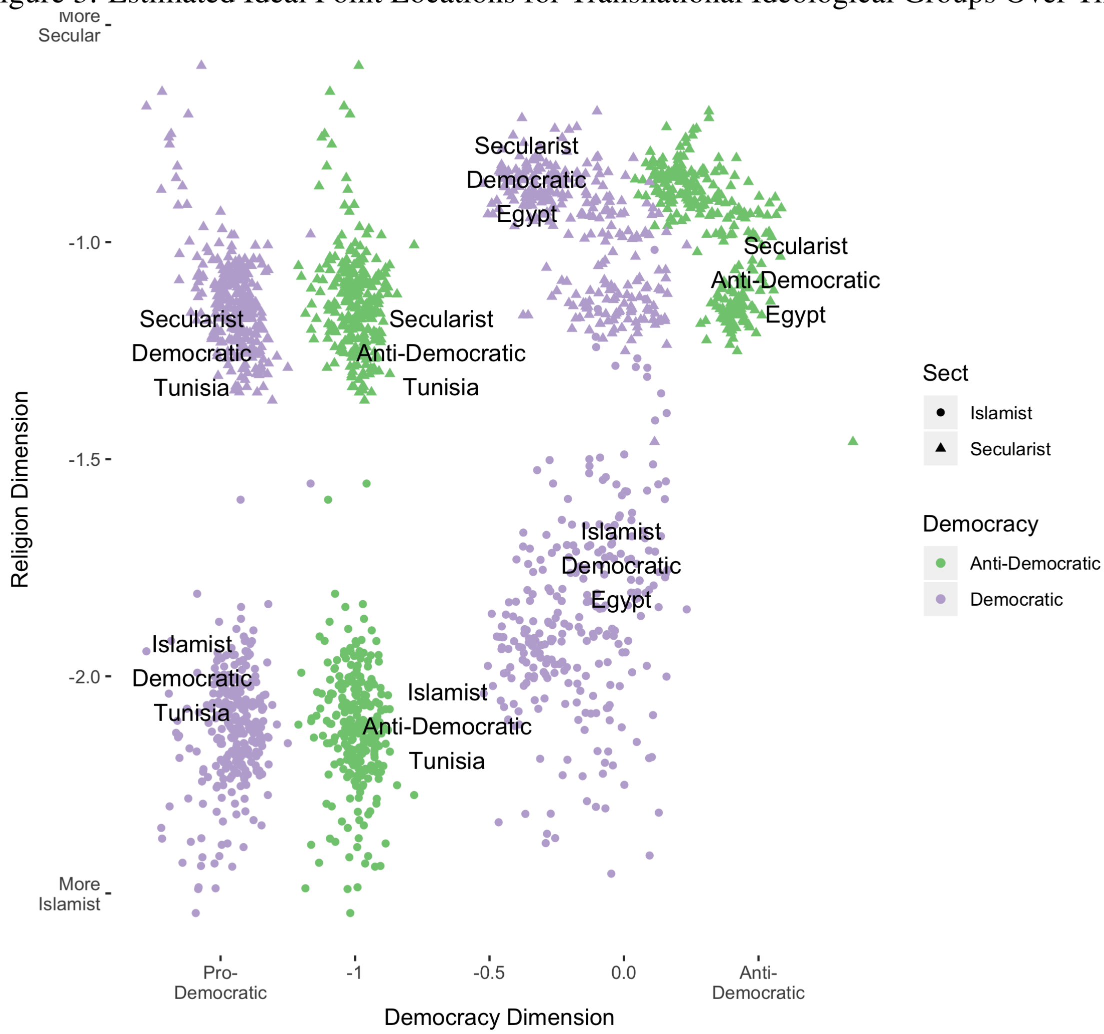
Estimating this model using full Bayesian inference would be computationally prohibitive given the volume of the data and the tens of thousands of model parameters. One possibility is to employ variational inference, which simplifies the joint posterior into factorized distributions that are much easier to sample (Kucukelbir et al. 2015). However, variational inference can understate uncertainty in its estimates, and the approximation can fail to adequately describe the true posterior (Grimmer et al. 2011; Kucukelbir et al. 2015). Instead we employed a novel technique recently developed by the Stan team for the parallelization of Markov chains. We used code that parallelized the gradient calculations necessary for each iteration of a Hamiltonian Markov sampler and are the primary computational burden. Employing a cluster computer system with 700 cores, we were able to estimate a converged chain within 48 hours.[9] Based on conversations with the Stan team in August 2019, we believe that this is the largest such model estimated to date. Our full Bayesian estimation produced 6,134 each of discrimination δ_j for all of the citizens in the model, but we will focus on the four group parameters that varied over time, $\alpha_{c,g,t}$, with one for each country-ideological pairing: Tunisian and Egyptian Islamists/secularists/democrats/authoritarians, as these are our primary focus for inference.

We first present a plot showing the estimated ideal points for the seven ideological groups in Figure 3. This chart plots each time point in the sample with a color and shape reflecting the two dimensions within the model: Islamist/secularist and democratic/authoritarian. While this flattened plot cannot show over-time trends, it can provide a clear description of the approximate locations of different groups in the model. The democratic Islamists in the lower right corner are the group with the largest over-time variance as evidenced by the spread in their ideal point locations. There is also more overall spread in the y axis (the religious dimension) versus the x axis (the democracy

9. To optimize computer time, we ran each chain for a total of 550 iterations, 300 of which were discarded as warmup iterations. While this number of iterations is smaller than other samplers, such as Gibbs samplers, it is sufficient for Hamiltonian Monte Carlo to reach convergence and exceeds the recommendations of the Stan manual (minimum 100 warmup iterations). All parameter estimates had Rhat values 1.1 or lower, and more than 99% of parameters had Rhat values lower than 1.05.

dimension), implying that sectarian identity is on the whole a stronger predictor of retweet patterns than political ideology during this time period.

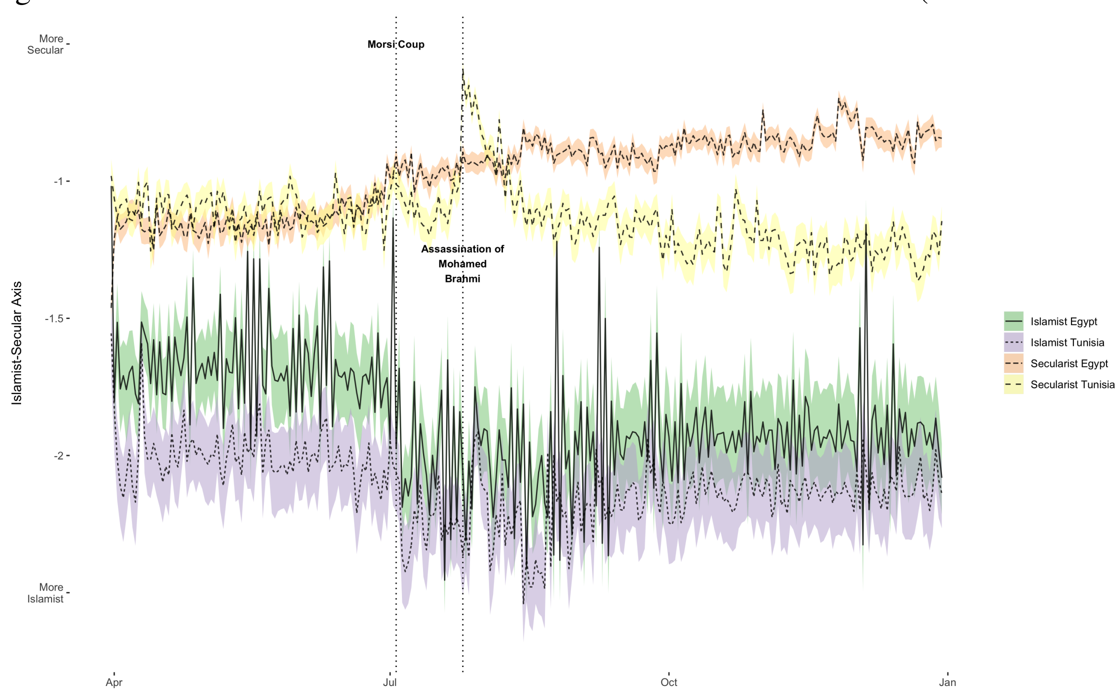
Figure 3: Estimated Ideal Point Locations for Transnational Ideological Groups Over Time



To examine the overall time trends, we plot each set of dimensions and groups separately in Figure 4 along with vertical lines showing when the following events occurred: 1) the military coup against Mohamed Morsi, and 2) the assassination of the secular leftist Tunisian politician Mohamed Brahmi. It should be noted that only Morsi's coup was explicitly parameterized in the model; the spikes in polarization occurring around the Brahmi's assassination emerged endogenously from the model. The confidence intervals on the chart reflects the 5%-95% quantiles of the empirical posterior, hereafter referred to as the high-posterior density (HPD) interval.

From this chart we can make useful descriptive inferences. The average position over-time of the time series gives us a sense of the general level of polarization in the data during this time period. First, the sectarian groups are much closer to each other than they are to the opposing group of the same country. These average locations would suggest that there is substantially ideological similarity in these groups as expressed in their retweet patterns. Second, polarization following Morsi's coup is easily evident on the chart without need to examine the model's coefficients. Islamists and secularists both diverged after the coup in opposite directions. Interestingly, while Islamists moderated somewhat over time, they remained closer than they were before the coup, while secularists start to diverge from each other after the coup. While some of these findings, such as the secularist divergence, invite speculation, the plot as a whole shows how it can capture many of the subtleties of this form of polarization graphically. Furthermore, the method is able to identify events that were particularly polarizing, especially the assassination of Mohamed Brahmi, without prior parameterization.

Figure 4: Over-Time Trends for Ideal Points of Secularists and Islamists (First Dimension)

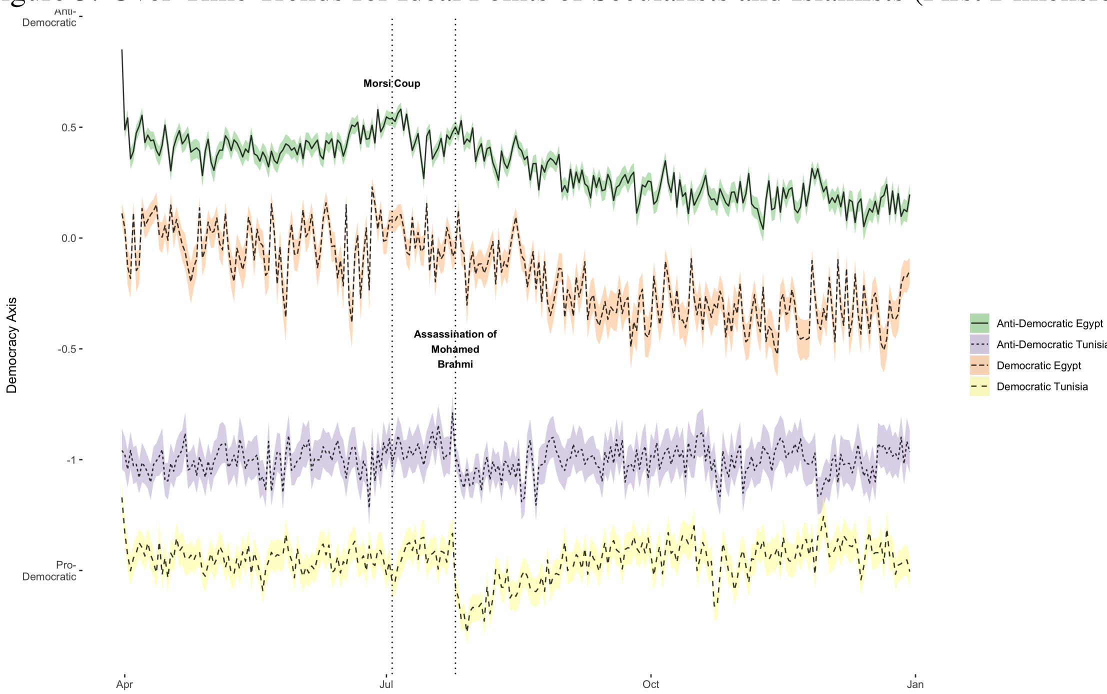


It is worthwhile to compare Figure 4 with Figure 1 that showed the raw counts of retweets for these same ideological groups. While Figure 1 had spikes around notable events like the coup against Morsi, the estimated ideal points in Figure 4 do not generally share these spikes and certainly not in the same proportions. The reason for this disparity is the work of the IRT-VAR measurement model, which will only pick up on changing proportions of tweets among users in terms of the pre-coded ideological groups. This disparity between raw data and the measured ideal points shows how the model is separating the chaff from the wheat by identifying salient trends in the data. For example, simple amplification of individual users, as happens due to Twitter "bots", will not affect the results of the model as it is the relative weight of ideological retweeting between groups that is being estimated.

We can similarly show the trends over time for our pro-democracy/anti-democracy dimension in Figure 5. This results are remarkably different than the trends for the religious dimension. First, we would note that the groups are more clustered in terms of their national versus sectarian identity. In other words, Tunisian and Egyptian democrats are more "democratic" than their country counter-parts, but Tunisians on the whole are more democratic than Egyptians. This scaling artifact should not be interpreted literally, but rather than the tweet patterns along this dimension are more distinct geographically than tweet patterns for the religious dimension. Interestingly, we do not see as much movement in general for Tunisians along this axis except for a pronounced spike following Brahmi's assassination. Egyptians, on the other hand, become more pro-democratic in their tweeting patterns over time.

Again, we need to be careful at interpreting these charts, but the over-time trends would suggest that both sides of the debate over democracy in Egypt were trying to reclaim the democracy space following the coup. This movement is likely due to the need to claim legitimacy arising from democratic discourse even as (or because of) democratic norms are violated. Indeed, even though the military regime in Egypt is seen as one of its most brutal in its history, the regime has faithfully implemented elections for the legislature and the executive, and has also held elections for referenda. As such, promoting pro-democratic discourse may have become paradoxically *more* important following the coup. We would note though that even as both groups in Egypt ex-

Figure 5: Over-Time Trends for Ideal Points of Secularists and Islamists (First Dimension)



hibit a pro-democratic shift, the net distance between them expressed in retweet patterns increased following the coup.

Following this descriptive analysis of the results, we turn to our inference on our parameters measuring the effect of Morsi's coup. To do so, we present an interpretation of the marginal effect of the coup on retweet counts in the first dimension (religion) in Figure 6 and the second dimension (democracy) in Figure 7. These marginal effects are based on those calculated by (Kubinec 2019) in which the effect of a parameter is weighted by the discrimination values, or the political positions of the citizens in this case. These marginal effects were created by averaging the exponentiated effect of the coup $\beta_{c,g,x}$ on the retweet count $Y_{c,g,t,j}$ over all of the positive and negative citizen discrimination parameters δ_j. These calculations are performed separately for each posterior draw $s \in S$ to capture uncertainty in the estimate:

$$\frac{\partial \hat{Y_{c,g,t,j}}}{\partial \beta_{c,g,x} | \delta_j > 0} = \frac{\sum_{s=1}^{S} \frac{\sum_{j=1}^{J_{\delta_j > 0}} e^{\beta_{c,g,x,s} \delta_{j,s}}}{J}}{S} \tag{5}$$

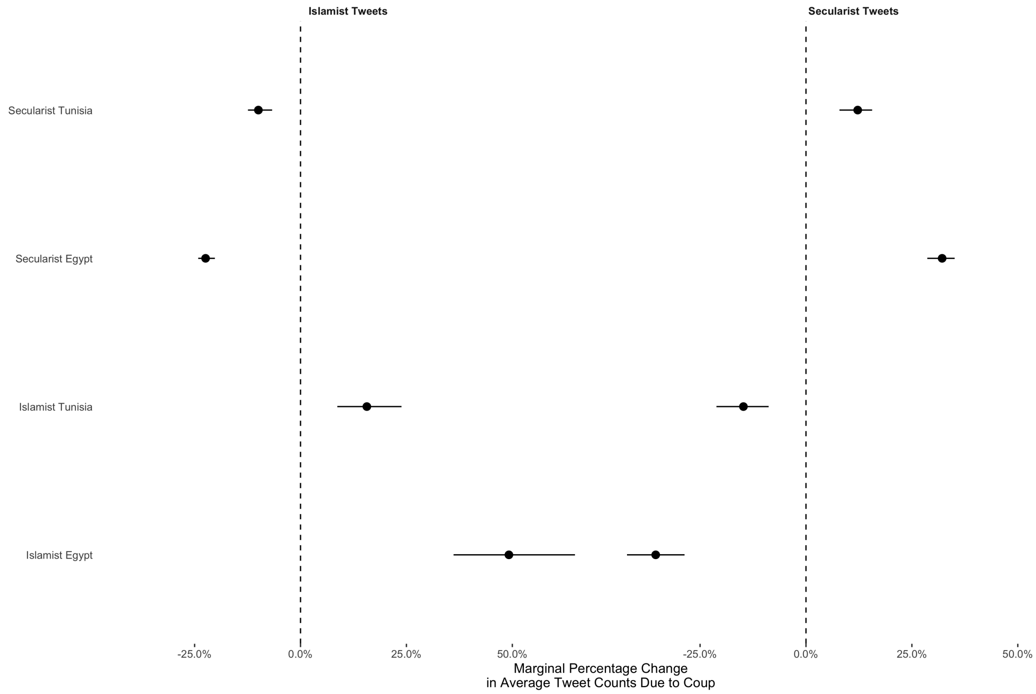
$$\frac{\partial \hat{Y_{c,g,t,j}}}{\partial \beta_{c,g,x} | \delta_j < 0} = \frac{\sum_{s=1}^{S} \frac{\sum_{j=1}^{J_{\delta_j < 0}} e^{\beta_{c,g,x,s} \delta_{j,s}}}{J}}{S} \tag{6}$$

These marginal effects are a way of interpreting the effect of the coup in terms of the observed retweet counts $Y_{c,g,t,j}$ that takes into account the total level of ideological polarization among Twitter users. The exponentation of the estimate allows us to interpret the coefficient as percentage change in retweet counts given that the outcome is modeled using the Poisson distribution. These percentages represent the increase or decrease in retweets coming from a typical Islamist or secularist user. In other words, as a result of weighting by the discrimination parameters, the estimate is *weighted by the underlying level of polarization among citizens*. For this reason, these ideal point marginal effects (Kubinec 2019) are a way to capture our full measurement uncertainty in the latent scale while providing a digestible "real world" interpretation. As can be seen in the Figures 6 and 7, there are two sets of marginal effects, one set for each end of the latent scale. An alternative way of phrasing the coefficients would be as the effect of the coup in encouraging retweets from ideologically like-minded and ideologically opposed citizens.

As can be seen in Figure 6, there were very strong effects of the coup on the tweet patterns of Islamist users. Elite Islamists in Egypt after the coup received 50% more retweets from Islamists and 25% fewer retweets from secularists. These effects are not identical because there are weighted by the total level of underlying polarization, or ideological identification, of these users. We see that the coup had the second largest effect on polarization among secularists in Egypt with similar increases in tweets from ideologically similar users. The effects are weaker for Tunisian groups, as might be expected given that the coup occurred in Egypt. In addition, the effects are much stronger for Tunisian Islamists than for secularists. Regardless of size, all of the effects point in a polarizing direction, e.g., more retweets coming from ideologically similar users and fewer from ideologically dis-similar users.

As might be expected from a perusal of the over-time trends for the second dimension, Figure 7 shows much weaker effects of the coup on ideological (de-)polarization for the democratic/authoritarian axis. First of all, we see that the effets of the coup are no longer uniformly

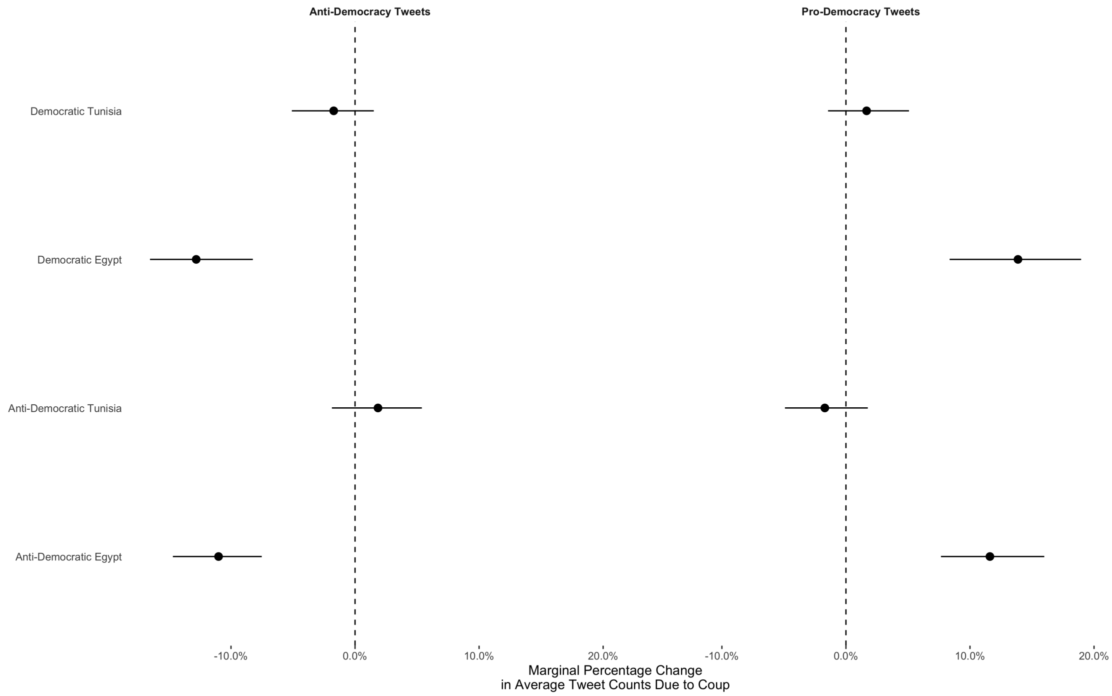
Figure 6: Marginal Effects of Coup Effect $\beta_{c,g,x}$ on Retweet Counts Weighted by Citizens' Religious Affiliation



polarizing. As could be seen visually in Figure 5, both Egyptian groups received more retweets from pro-democratic users and decreased retweets from anti-democratic users. The effects of the coup on Tunisian groups, on other hand, were too weak to be statistically distinguishable. The model provides only limited evidence that authoritarian Tunisians became more authoritarian and pro-democratic Tunisians became modestly more pro-democratic, but the opposite conclusion is still plausible. On the whole, these figures reveal that most of the polarization occurring as a result of the coup is along the religious identity axis, although there is interesting tandem movement in the pro-democracy direction for Egyptian groups as we described earlier based on visual analysis.

Finally, we can examine whether or not there was additional influence on the ideological locations of groups due to *transnational* influence. To do so, we calculate the impulse-response functions for a ten-day window following the coup. As explained previously, these impulse response functions (IRFs) show the average over-time decay of the effect of the coup on the ideal points of different groups. By altering the parameters of the simulation used to calculate these

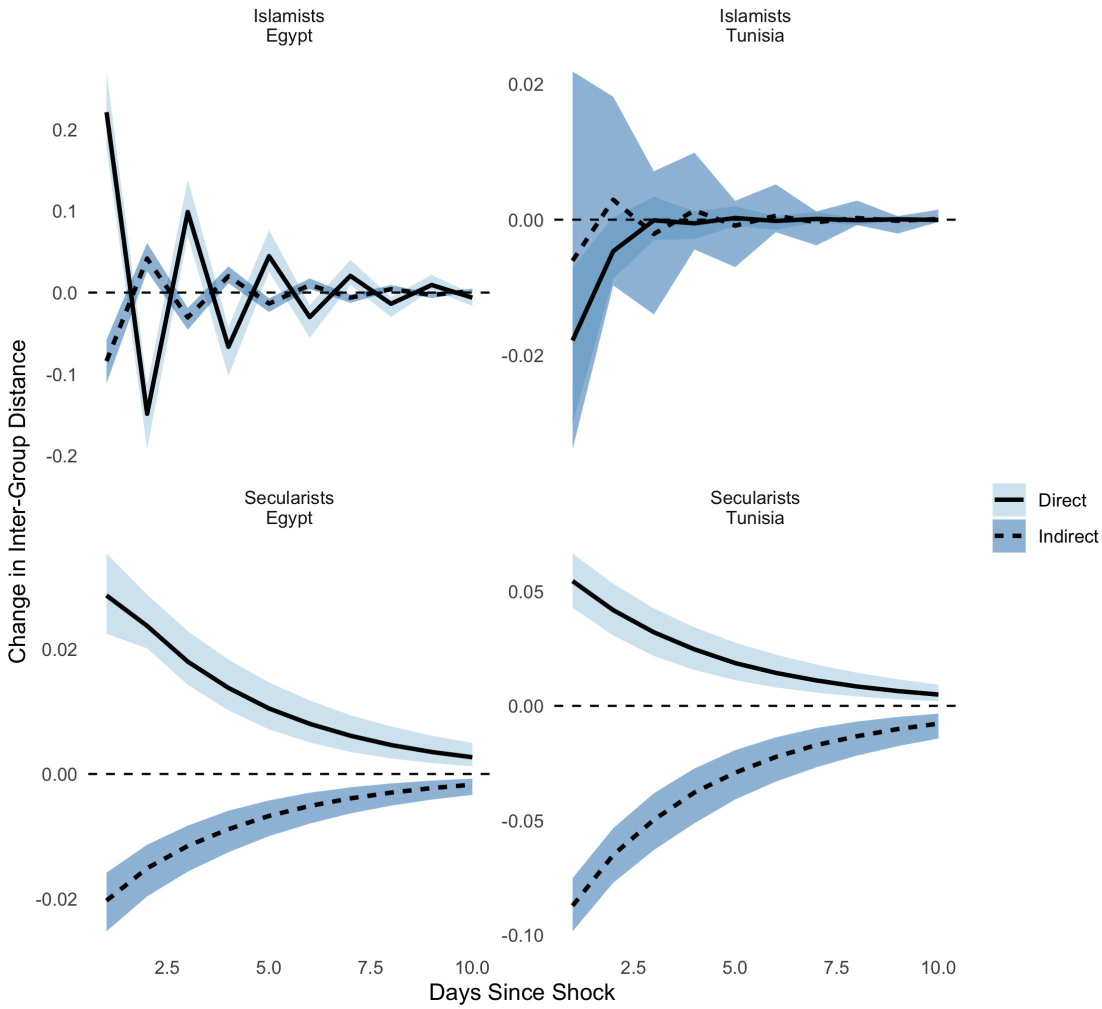
Figure 7: Marginal Effects of Coup Effect $\beta_{c,g,x}$ on Retweet Counts Weighted by Citizens' Pro-Democratic Inclinations



effects, we can separately identify the *direct* effect of the coup on a group and the *indirect* effect of the coup on a group's transnational allies traveling back to influence its own political position. Doing so is straightforward as it involves setting one of the autoregressive lag coefficients $\beta_{c,g,t,I}$ for internal influence or $\beta_{c,g,t,E}$ for external influence equal to zero. The results of these direct (light blue) and indirect (dark blue) effects for the religious dimension are shown in Figure 8, while the same IRFs for the democracy dimension are shown in Figure . The scale in these figures is taken from movement in the underlying ideal point scale.

What is immediately apparent from these figures is that for three out of the four groups, the indirect and direct effects are mirror images of each other. Structurally, the Egyptian Islamists show a different type of over-time autocorrelation as their ideal points oscillate back and forth over time. This behavior is a result of a negative autoregressive parameter $\beta_{c,g,t,I}$ that allows the series to oscillate. It is difficult to ascribe a clear meaning to this oscillation, although it should be noted that the Egyptian Islamists exhibit the largest amount of over-time change (i.e., variance) relative

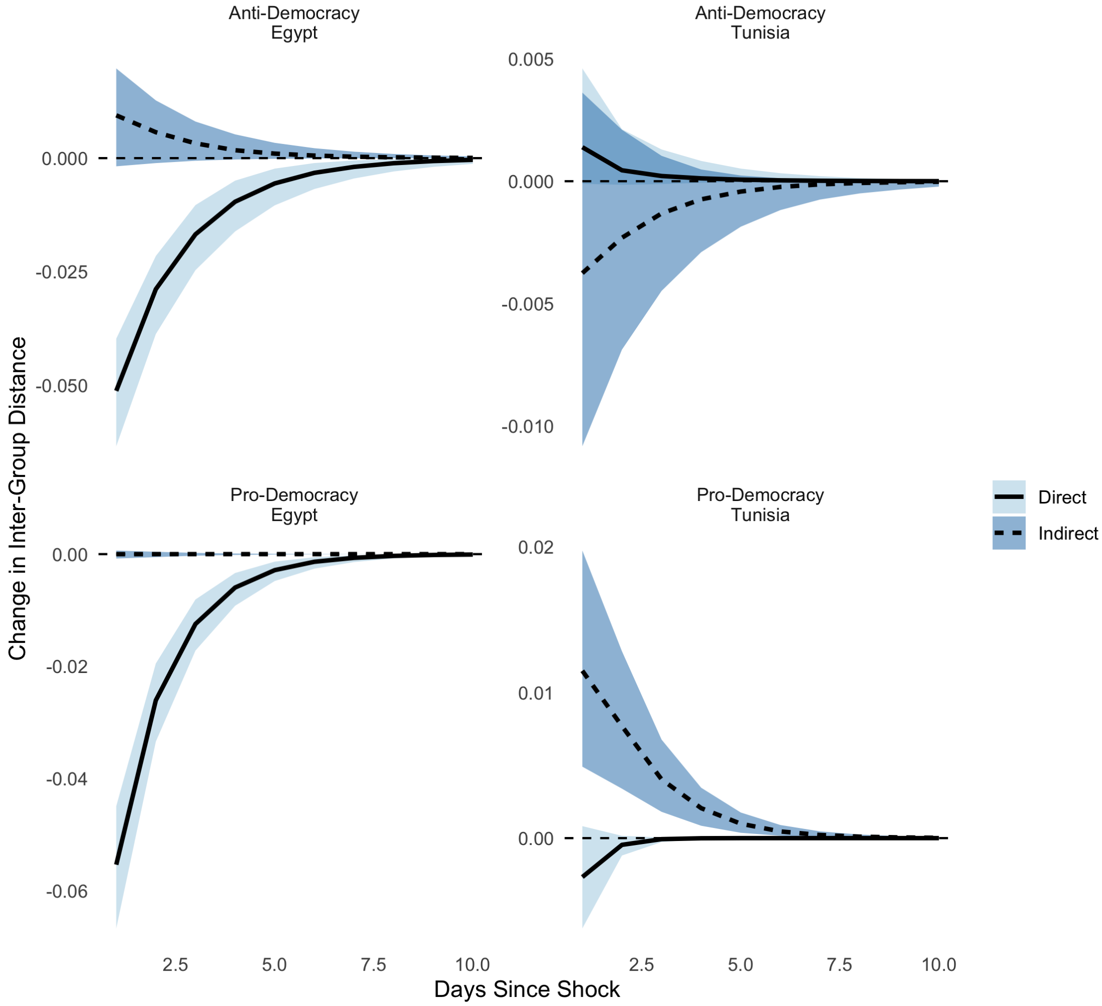
Figure 8: Indirect and Direct Effects of Coup Over Time on Islamists and Secularists



to other groups. It could be a sign of the stress that the group is under during this time.

However, whether the time series exhibit oscillation or stable over-time decay, the indirect effects counter-balance the direct effects. What this result implies is that these ideological groups were *de-polarized* by the polarization occurring among their transnational allies. The effects are strongest for the Egyptian Islamists. As their counterparts in Tunisia polarized due to the coup, Egyptian Islamists de-polarized. This inference is made possible because the model permits us to examine crucial unobserved counterfactuals: if the coup had occurred but the Tunisian Islamists had not existed, what would have been the effect of the coup on Egyptian Islamists? The answer

Figure 9: Indirect and Direct Effects of Coup Over Time on Democrats and Authoritarians



is that Egyptian Islamists would have ended up *even more* polarized than they did in fact become. The pattern is similar for secularists in both countries, though the effects are relatively weaker.

Fascinatingly, Tunisian Islamists show little indirect influence coming from their ideological allies. This minimal influence occurred because the indirect influence parameter for the Tunisian Islamists is estimated very close to zero with a posterior mean of 0.017 (HPD -0.088, 0.112). While the Tunisian Islamists reacted to the *coup* very strongly, they did not react at all to the polarization occurring between their transnational allies and secularist adversaries. Again, it is the model that permits us to test this important counterfactual in a highly endogenous system: how would the

Tunisian Islamists have reacted if the Egyptian Islamists had not existed? As it turns out, not very differently at all.

Finally we turn to the similar effects for the democracy dimension in Figure 9. Just as the effect of the coup was weaker for this dimension, so the indirect/direct simulations show less information. The only pronounced indirect effect is for Tunisian democrats, who became slightly more authoritarian (positive) due to the movement of their ideological allies. This finding suggests that there is indeed some kind of strategic positioning at work in democratic discourse among democrats and authoritarians following the coup with the definition of democracy itself a contested value.

Discussion

The substantive interpretation of the movement of group-level ideal points presented in the previous section will require further theoretical development. Without having such precise measures of exactly how groups respond to each other in real time, we did not previously have the ability to come up with hypotheses that describe how Islamists and secularists will react to each other on a day by day basis. Furthermore, explanation of the exact trajectory of the ideal points over time will have to involve further specification of the sometimes conflicting determinants of group polarization that we presented in this paper. For example, in Tunisia, Islamists were under considerable pressure to dampen ideological conflict during this time period as they faced rising social unrest due to Islamist radical violence (McCarthy 2016). For that reason, after the coup in Egypt they may have feared supporting their co-religionists too publicly, lest they suffer a similar fate within their own country (Grewal 2016). This kind of suppression, or what we call de-polarization, may explain why Egyptian Islamists de-polarized in response to witnessing polarization of Tunisian Islamists and secularists.

In addition, the limited influence of Egyptian Islamists on Tunisian Islamists may be explained by the finding reported by Marks (Marks 2015) that Tunisian Islamists, in interviews in 2011, did not identify strongly with the Egyptian Muslim Brotherhood, but instead saw themselves as similar to Turkey's Justice and Development Party, a more moderate and modernizing Islamist party. While the two parties share many ideological similarities, the particularities of how they perceive themselves in the wider world affect their risk of exposure to transnational ideological polarization. We believe that our model's ability to identify these movements at a low level of granularity will provide fodder for further theoretical development of how group-level polarization actually occurs. As with modeling of other complex dynamic systems like weather patterns, we do not expect predictions to be deterministic, yet there is still much more we could learn empirically from this model.

For our data, we are limited in modeling dynamics before 2012 because many of the Islamists did not open Twitter accounts until the beginning of 2013, leaving limited ideological diversity prior to that date. This limited time series may explain why we do not observe an even larger polarizing effect of the anti-Morsi coup: by the beginning of 2013, groups had already become polarized, which would leave less room for yet more polarization to occur. Worth (2016, 133-169) describes how, over the course of the Morsi presidency, a prominent moderate Egyptian Muslim Brotherhood politician, Mohamed Beltagy, became progressively alienated from secularists with whom he had previously cooperated. The story of Beltagy may well point to a more general phenomenon that depresses our polarization result.

Yet even with these caveats, our findings provide important insight into the underlying dynamics of transnational ideological polarization. It is difficult to study endogenous processes whether qualitatively or quantitatively, and we are able to analyze a very large dataset and reduce it down to the particular latent social cleavage of interest. Given how responsive the inter-group distance measure is to clearly polarizing events–some of which we did not identify in advance, such as the assassination of the Tunisian secularist politician Mohammed Brahmi–our model's internal validity appears quite high.

Conclusion

For all of the theoretical and empirical progress in the study of ideological diffusion, lacking have been studies demonstrating the spread of contention across national boundaries that account for pre-existing contention in the receiving country. In this paper we address this gap by putting forward a method of estimating the latent positions of ideological groups in Egypt and Tunisia during the tumultuous period of the Arab Uprising. We take advantage of Twitter's widespread usage to measure transnational polarization by showing how latent ideological scores change over time and also in tandem with similar ideological groups in other countries. The use of the IRT-VAR model allows for these estimates to incorporate measurement uncertainty while also providing useful summary measures of inter-group distance.

The precision of the hypothesis tests we are able to implement in this paper enables us to identify the direct and indirect effects of group polarization. Being able to separate these different components of the feedback process allows us to substantiate the major elements of the theory, and to support the central point of our paper that transnational linkages among ideological groups can endogenously heighten or dampen polarization independent of what is occurring within each group's country. While we are not the first to document such linkages, we are the first to directly measure this kind of transnational ideological polarization in a way that incorporates our uncertainty in measuring latent social cleavages.

Of course, the spread of group polarization is only one narrow question in the general question of international diffusion. Mechanisms, the efficacy of diffusing and blocking agents, the diffusing phenomenon itself, and many other factors are needed to complete the picture. It is worth noting that thus far the political outcome in Tunisia – enduring if fragile democracy – is different from that in Egypt – a military-authoritarian secular regime. Still, it is important to establish empirically whether and through what mechanisms contentious politics can diffuse across national borders, and that is what we have done. Our findings open the door to further exploration of the determinants and measurement of transnational polarization during periods of political instability. This model provides a rich range of estimates and can pinpoint places at which polarizing events occurred.

Furthermore, we can show how short-term shocks translate into long-term differences in polarization over time. We hope that this evidence stimulates more investigation of the determinants and effects of group feedback effects in ideologically polarized societies.

References

Aday, Sean, Henry Farrell, Marc Lynch, John Sides, and Deen Freelon. 2012. *New Media and Conflict after the Arab Spring.* Peaceworks. U. S. Institute of Peace. Accessed July 25, 2018. https://www.files.ethz.ch/isn/150696/PW80.pdf.

Barberá, Pablo. 2015. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23:76–91.

Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58 (2): 367–386.

Box-Steffensmeier, Janet M., John R. Freeman, Matthew P. Hitt, and John C. W. Pevehouse. 2014. *Time Series Analysis for the Social Sciences.* Cambridge: Cambridge University Press.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Rollcall Data." *American Political Science Review* 98 (2): 355–370.

Conover, M. D., J. Ratkiewicz, M. Francisco, B. Goncalves, A. Flammini, and F. Menczer. 2011. "Political Polarization on Twitter," *Association for the Advancement of Artificial Intelligence.*

Coppedge, Michael, John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, et al. 2017. *V-DEM Dataset v7.* Working Paper. Social Science Research Network, May 15. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2968289.

Coser, Lewis A. 1956. *The Functions of Social Conflict.* Routledge.

Ellemers, Naomi, Paulien Kortekaas, and Jaap W. Ouwerkerk. 1999. "Self-categorisation, Commitment to the Group and Group Self-Esteem as Related But Distinct Aspects of Social Identity." *European Journal of Social Psychology* 29 (2): 371–389.

Franzese Jr, Robert J., and Jude C. Hays. 2008. "Independence in Comparative Politics: Substance, Theory, Empirics, Substance." *Comparative Political Studies* 41 (4): 742–780.

Freelon, Deen, Marc Lynch, and Sean Aday. 2015. "Online Fragmentation in Wartime: A Longitudinal Analysis of Tweets about Syria, 2011-2013." *The Annals of the American Academy of Political and Social Science* 659:166–179.

Freeman, John R., John T. Williams, and Tse-min Lin. 1989. "Vector Autoregression and the Study of Politics." *American Journal of Political Science* 33 (4): 842–877.

Gerring, John. 2012. *Social Science Methodology: A Unified Framework.* Second. New York: Cambridge University Press.

Grewal, Sharan. 2016. "A Quiet Revolution: The Tunisian Military after Ben Ali." *Carnegie Middle East Center.* http://carnegie-mec.org/2016/02/24/quiet-revolution-tunisian-military-after-ben-ali/iucy.

Grimmer, Justin, Eitan Hersh, Brian Feinstein, and Daniel Carpenter. 2011. "Are Close Elections Random?" Working Paper, Stanford University, January.

Jamal, Amaney A., Robert O. Keohane, David Romney, and Dustin Tingley. 2015. "Anti-Americanism and Anti-Interventionism in Arabic Twitter Discourses." *Perspectives on Politics* 13 (1): 55–73.

Kropko, Jonathan. 2013. *Measuring the Dynamics of Political Power: a Time-Series IRT Model.* Presentation. Political Methodology Society.

Kubinec, Robert. 2017. *Absence Makes the Ideal Points Sharper.* Poster. Political Methodology Society.

Kubinec, Robert. 2019. "Generalized Ideal Point Models for Time-Varying and Missing-Data Inference." *Open Science Foundation Preprints.* doi:10.31219/osf.io/8j2bt.

Kucukelbir, Alp, Rajesh Ranganath, Andrew Gelman, and David Blei. 2015. "Automatic Variational Inference in Stan." In *Advances in Neural Information Processing Systems 28,* edited by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, 568–576. Curran Associates, Inc. http://papers.nips.cc/paper/5758-automatic-variational-inference-in-stan.pdf.

Marks, Monica. 2015. *Tunisia's Ennahda: Rethinking Islamism in the Context of ISIS and the Egyptian Coup.* Working Paper. The Brookings Institution, August. https://www.brookings.edu/wp-content/uploads/2016/07/Tunisia_Marks-FINALE-5.pdf.

Martin, Andrew D., and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999." *Political Analysis* 10 (2): 134–153.

McCarthy, Rory. 2016. "The Tunisian Uprising, Ennahda and the Revival of an Arab-Islamic Identity." In *Political Identities and Popular Uprisings in the Middle East,* edited by Shabnam J. Holliday and Philip Leech, 157–176.

Melamed, Yitzhak Y. 2013. *Spinoza's Metaphysics: Substance and Thought.* Oxford University Press.

Mercer, Jonathan. 1995. "Anarchy and Identity." *International Organization* 49 (2): 229–252.

Mitts, Tamar. 2019. "From Isolation to Radicalization: Anti-Muslim Hostility and Support for ISIS in the West." *American Political Science Review,* 113th ser., 1:173–194.

Moulaert, Frank, and Farid Sekia. 2003. "Territorial Innovation Models: A Critical Survey." *Regional Studies* 37 (3): 289–302.

Owen, John M., and William Inboden. 2015. "Putin, Ukraine and the Question of Realism." *The Hedgehog Review* 17 (1): 86–97.

Quine, W. V. 1989. *Quiddities: An Intermittently Philosophical Dictionary.* Harvard University Press.

Sambanis, Nicholas, and Moses Shayo. 2013. "Social identification and ethnic conflict." *American Political Science Review* 107 (2). http://journals.cambridge.org/repo_A895v1Rl.

Siegel, Alexandra A., Joshua A. Tucker, Jonathan Nagler, and Richard Bonneau. 2018. *Tweeting Beyond Tahrir: Ideological Diversity and Political Intolerance in Egyptian Twitter Networks.* Working Paper. New York University, March. https://alexandra-siegel.com/wp-content/uploads/2018/03/Egypt_Tolerance_March2018.pdf.

Simmel, Georg. 1898. "The Persistence of Social Groups." *American Journal of Sociology* 3 (5): 662–698.

———. 1955. "A Contribution to the Sociology of Religion." *American Journal of Sociology* 60 (6): 1–18.

Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica* 48 (1): 1–48.

Smith, John. 2012. "Reputation, Social Identity and Social Conflict." *Journal of Public Economic Theory* 14 (4): 677–709.

Solingen, Etel. 2012. "Of Dominoes and Firewalls: The Domestic, Regional, and Global Politics of International Diffusion." *International Studies Quarterly* 56 (4): 631–644.

Steinert-Threlkeld, Zachary C. 2017. "Spontaneous Collective Action: Peripheral Mobilization During the Arab Spring." *American Political Science Review* 111 (2): 379–403.

Takane, Yoshio, and Jan de Leeuw. 1986. "On the Relationship Between Item Response Theory and Factor Analysis of Discretized Variables." *Psychometrika* 52 (3): 393–408.

Taylor, Paul W. 1975. *Principles of Ethics: An Introduction.* Dickenson Publishing Company.

Tilly, Charles. 2005. *Trust and Rule.* Cambridge University Press.

Treier, Shawn, and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52 (1): 201–217.

Weber, Ingmar, Venkata R. Kiran Garimella, and Alaa Batayneh. 2013. "Secular vs. Islamist Polarization in Egypt on Twitter." In *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).* August 25.

Worth, Robert. 2016. *A Rage for Order: The Middle East in Turmoil, from Tahrir Square to ISIS.* New York: Farrar, Straus / Giroux.

Zhukov, Yuri M., and Brandon Stewart. 2013. "Choosing Your Neighbors: Networks of Diffusion in International Relations." *International Studies Quarterly* 57 (2): 271–287.

Zivot, Eric, and Jiahui Wang. 2006. *Modeling Financial Time Series with S-PLUS.* Springer.