Rotman & Shalev (SMR)
Online Appendices

Appendix 1

Advance preparation and post-processing of a mobile location dataset

Advance preparation refers to the researcher's responsibility to supply the data provider with a precise description of each protest event for which data are sought. Specifically, the provider requires both time-and-date information and geographical polygons. In the case of an event that has not yet occurred, the researcher should develop a plan for field observation and arrangements to comprehensively document the demonstration on visual media that can be analyzed after the event. Video recordings made by drones could be helpful in this respect, not only to assist in defining the time-space dimensions of protest events but potentially also as a basis for independently estimating crowd size.[1]

Turning to post-processing, it should be borne in mind that the procedures which will now be described were adopted specifically for the data described and analyzed in this paper. As explained in Part 1, the provider delivered a dataset with estimates of the aggregate number of persons present at selected events and their distribution across imputed localities of residence. This dataset included benchmark observations based on the number present at the location and hours of each protest event a week before it occurred. The main task in preparing the data for analysis was to determine how to use these benchmarks for adjusting the gross estimates of protest participation at the time and place of each event. Since not all people caught in the area of a protest event are demonstrators, the gross number of observed persons must be downwardly adjusted to account for bystanders. While the benchmarks provide an indication of the number of people normally present in the area of measurement, this number is most likely higher than the (unobservable) number of bystanders, since during a mass demonstration

[1] Cf. Choi-Fitzpatrick, Austin. 2015. "Drones, Data, and Tactics." *Mobilizing Ideas*, https://mobilizingideas.wordpress.com/2015/03/09/more-measurements-more-diverse-sources-for-studying-mobilization-yes-please, accessed 15 August 2017

normal activities are likely to be at least partially curtailed. Accordingly, the benchmark estimates need to be deflated before subtracting them from the gross estimates of protest participation.

Several considerations were taken into account in computing the deflation factor used for each combination of event and home locality.

1. The likelihood of active participation by residents of each home locality.
   Persons from locations whose residents are often observed at demonstrations over the course of a campaign are more likely to be demonstrators. Thus, when a locality has a high rate of overall protest participation, its benchmark is deflated to a greater extent and consequently a larger share of the initial estimate of protest participants is defined as active demonstrators. The likelihood of active participation was estimated using a composite measure (generated using factor analysis) of three indicators of strong engagement in the campaign: (1) the ratio between the highest protest count throughout the campaign and the size of the working-age population, (2) the ratio between the average protest count and the size of the working-age population, and (3) the ratio between the average protest count and the average benchmark observation.
2. The distance of each home locality from the host locality where a demonstration took place.
   Given the effort involved, persons who come from afar are more likely to be demonstrators than mere passersby, whereas persons that live in the vicinity of a protest are more likely to be in the area for reasons unrelated to the protest. Thus, when distance from the protest location is high, the benchmark was deflated to a greater extent. The natural log of the distance[2] was used for this purpose, on the

[2] To avoid negative values after the logarithmic transformation, a value of 1 was added to all distances before the transformation. The transformed variable ranges between 0 and 6.04.

assumption that small differences close to the event's location have a larger effect than the same differences further away from the demonstration.

3. The likelihood that the area of a demonstration hosts other kinds of activities. To the extent that activities other than protesting are unlikely to take place, the benchmark is more deflated. This was estimated based on (1) the size of the measurement area (larger size is expected to raise the chances of other kinds of activities), (2) the presence of facilities for commercial and leisure activities in the measurement area (offering more opportunities for non-protest activities), and (3) the accessibility of the area, as indicated by bus stops and parking areas (non-protest activities are more likely in more accessible areas). Data on these properties of the receiving locations were gathered and then rated independently by two judges on a 5 point scale (with a correlation of 0.7 between the two raters). The average of the two scores was used for calculating the deflation factor.

In applying these three considerations it is important to distinguish between two categories of persons who may be observed at protest events – residents of the area in which the protest took place, and all others ("outsiders"). In practice the benchmark for residents was only deflated by the first criterion, based on their overall rate of protest participation over the course of the campaign. Because the distance of residents from the protest event is equal to zero, the second criterion would have no effect on their benchmark and was not applied. The third criterion was also not applied to residents, on the grounds that because they live in the area of the demonstration they could always have other reasons for being there.

A general problem in implementing the three criteria for benchmark deflation is that there is no way of determining their relative importance. Consequently, multiple versions of the deflation factor were calculated (see Eq. 1 below). Three possible weights were assigned to each indicator, resulting in 27 different combinations for non-residents but only 3 for residents, since only the first indicator was used for calculating their deflation factors. The highest score for the deflation factors in our data matrix is 1,

and the lowest ranges between 0.025 using the highest weights to 0.675 using the lowest. After multiplying the benchmark reading by the deflation factor, it was subtracted from the gross count of the number present (see Eq. 2 below). This yielded 27 versions of the number of participations from each Statistical Area at each event. After bottom coding (replacing negative values with zeros), the median of the 27 versions of the adjusted number of participants was used for further analysis.

Note that participation by residents cannot be accurately measured by this procedure. It can only identify an unusually high presence of residents in the area at the time of a protest event. Due to their regular presence at event locations, residents inevitably have high benchmark observations, and there is no way of knowing from the data whether at the time of the protest they stayed at home, joined the protest, or went out to observe it. As a result, our calculation of residents' participation is almost certainly conservative.

Eq. 1. Deflation factors.

Visitors: $DF = 1 - (W_1 * Active_i) - (W_2 * \ln(Distance + 1)_{ij}) - (W_3 * OtherActs_j)$

Residents: $DF = 1 - (W_1 * Active_i)$

DF is the deflation factor; *Active* is a measure of likelihood of active participation from Statistical Area *i*; *Distance* is the distance from Statistical Area *i* to event *j*; *OtherActs* is a measure of likelihood of other kinds of activities to take place in the area of event *j*; W_1 is equal to either 0.01, 0.02 or 0.03; W_2 is equal to either 0.01667, 0.033 or 0.05; W_3 is equal to either 0.025, 0.05 or 0.075

Eq. 2. Net protest participation.

$NetMax_{ij} = GrossMax_{ij} - (Benchmark_{ij} * DF_{ij})$

NetMax is the highest adjusted number of participants from Statistical Area *i* observed at event *j*; *GrossMax* is the highest gross number of persons from Statistical Area *i* observed during event *j*; *Benchmark* is the average number of persons from Statistical Area *i* observed in the area of event *j* a week before the demonstration; *DF* is the deflation factor used to reduce the size of the benchmark.

Appendix 2

Sensitivity Tests

As detailed in Appendix 1, in order to obtain estimates of the number of protesters net of presumed bystanders, counts of the number of cellular devices observed in the catchment area of a demonstration have been deflated in order to account for bystanders. The estimates of number of participants (net of bystanders) that served as the basis for the analysis presented in the paper are the median estimations out of 27 alternative values that vary in the scheme used to deflate the raw number of devices observed at each event. Since the choice of the median could be seen as arbitrary, to test the robustness of the analysis sensitivity tests were carried out comparing results based on the median estimations to those based on the minimum and maximum estimations (i.e. the most and least deflated).

As explained in the text, the primary advantage of our locational participation data lies in the ability to examine relationships between the sociodemographic characteristics of home localities and their rates of protest participation. Therefore, we examine to what extent such relationships are sensitive to the degree of deflation. The results indicate that the relationships of interest remain highly similar when varying the deflation scheme, leading us to conclude that the exact variant of the formula used to account for bystanders has only trivial effects on the findings.

Table A2.1 presents the mean and standard deviation of the participation rate among the majority population and the three minority sectors (weighted by locality size), along with ratios between minority and majority means. No substantial differences are observed between the different estimates of participation rates. Focusing on the majority population, the correlations between participation and both the share of college graduates in the adult population and the share of votes for left and center parties in the 2009 elections are presented in Table A2.2. Again, no substantial differences can be observed.

Table A2.3 shows results from multivariate logistic regressions that estimate the effects of the share of votes for left/center parties, share of college graduates, distance to nearest event, and population sector on a home locality's probability of inactivity (participation rate below 1%). Tables A2.4 reports effects of the same IVs on the logged peak participation rate of active localities, using multiple fractional polynomial GLM. Three versions of each dependent variable were modeled, based on the minimum, median and maximum participants estimations. Comparing the coefficients of the independent variables in both tables reveals that they hardly differ between the models that utilize alternative versions of the participation estimates.

The results presented here reassure that the relationships observed in the data are not sensitive to how the initial values are adjusted to account for the presence of bystanders. The deflation scheme proposed in this paper, while plausible, lacks external sources that can validate it. Nevertheless, it seems that the meaningful conclusions from analyzing the data are independent of the exact degree of deflation employed.

Table A2.1

Peak participation rates by population sector and benchmark deflation

Sector	***Deflation***	Low	Medium	High
Majority	Mean	12.5%	12.8%	13.0%
	Std. Dev.	0.15	0.15	0.16
Arabs	Mean	6.9%	7.0%	7.2%
	Std. Dev.	0.11	0.11	0.12
	Mean relative to majority	0.55	0.55	0.55
Ultra-orthodox	Mean	4.2%	4.1%	4.4%
	Std. Dev.	0.02	0.02	0.02
	Mean relative to majority	0.33	0.32	0.34
Settlers	Mean	6.5%	6.5%	6.5%
	Std. Dev.	0.08	0.08	0.08
	Mean relative to majority	0.52	0.51	0.50

Table A2.2

Spearman correlations between peak participation and sociopolitical characteristics (majority localities only), by benchmark deflation

	Low	Medium	High
Share of college graduates	0.34	0.33	0.32
Share of votes to left/center	0.43	0.42	0.41

Table A2.3

Logistic models predicting probability of inactivity, by benchmark deflation

	Low	Medium	High
%left/center	-0.004	-0.004	-0.004
	(.00)	(.00)	(.00)
%college	-0.02	-0.02	-0.02
	(.00)	(.00)	(.00)
Distance	0.07	0.07	0.07
	(.01)	(.01)	(.01)
Arabs	0.78	0.78	0.79
	(.17)	(.17)	(.17)
Ultra-orthodox	0.55	0.45	0.45
	(.23)	(.23)	(.23)
Settlers	0.50	0.50	0.51
	(.27)	(.27)	(.27)
Constant	-0.42	-0.45	-0.45
	(.14)	(.14)	(.14)

Table A2.4

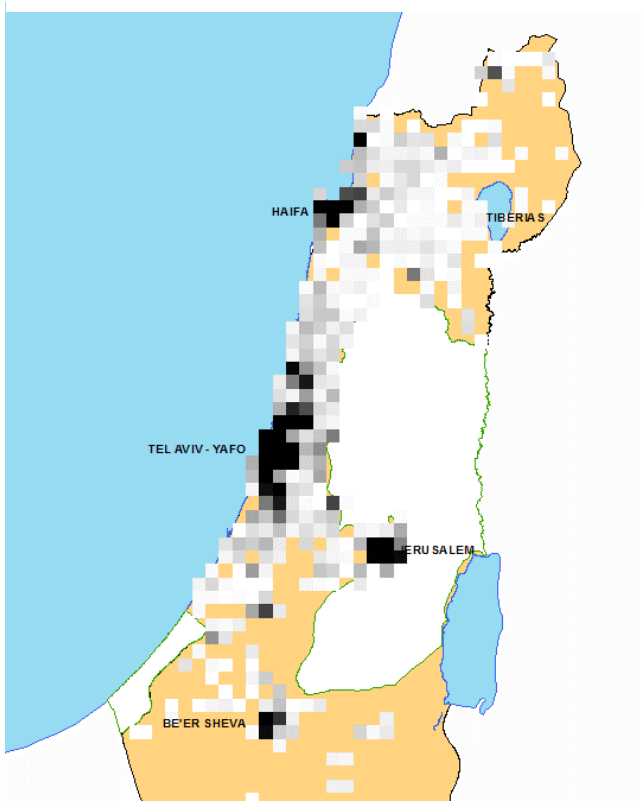
Multiple fractional polynomial GLM models predicting logged peak participation rate, by benchmark deflation

	Low	Medium	High
%left/center (power 0.5)	-1.48	-1.46	-1.46
	(.22)	(.22)	(.22)
%left/center	0.56	0.55	0.55
	(.06)	(.06)	(.06)
%college	0.01	0.01	0.01
	(.00)	(.00)	(.00)
Distance (power -1)	4.01E-07	4.54E-07	5.20E-07
	(.00)	(.00)	(.00)
Arabs	-0.48	-0.49	-0.49
	(.17)	(.18)	(.18)
Ultra-orthodox	-0.92	-0.95	-0.92
	(.20)	(.19)	(.19)
Settlers	-0.25	-0.23	-0.23
	(.17)	(.17)	(.17)
Constant	-2.67	-2.67	-2.66
	(.04)	(.04)	(.04)

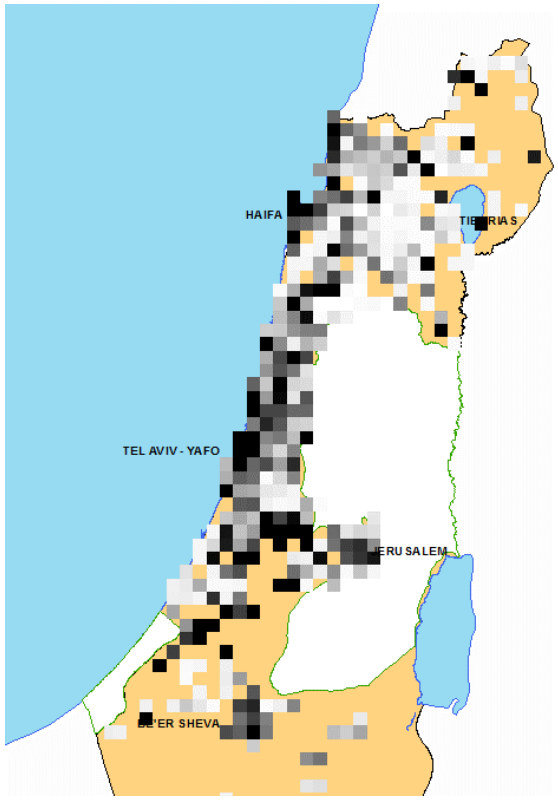
Appendix 3

Geographical dispersion of the protest

Map A3.1: **Distribution of total participations between home localities**



Map A3.2: **Participation rates per home locality**



Map A3.3: **Location of demonstrations** (bubbles show number of participants)

Notes: In **Maps A6.1 and A6.2** participation is aggregated in a grid comprising 20km*20km squares. The amount or intensity of participation are represented by a gray scale, with darker tones indicating higher levels.

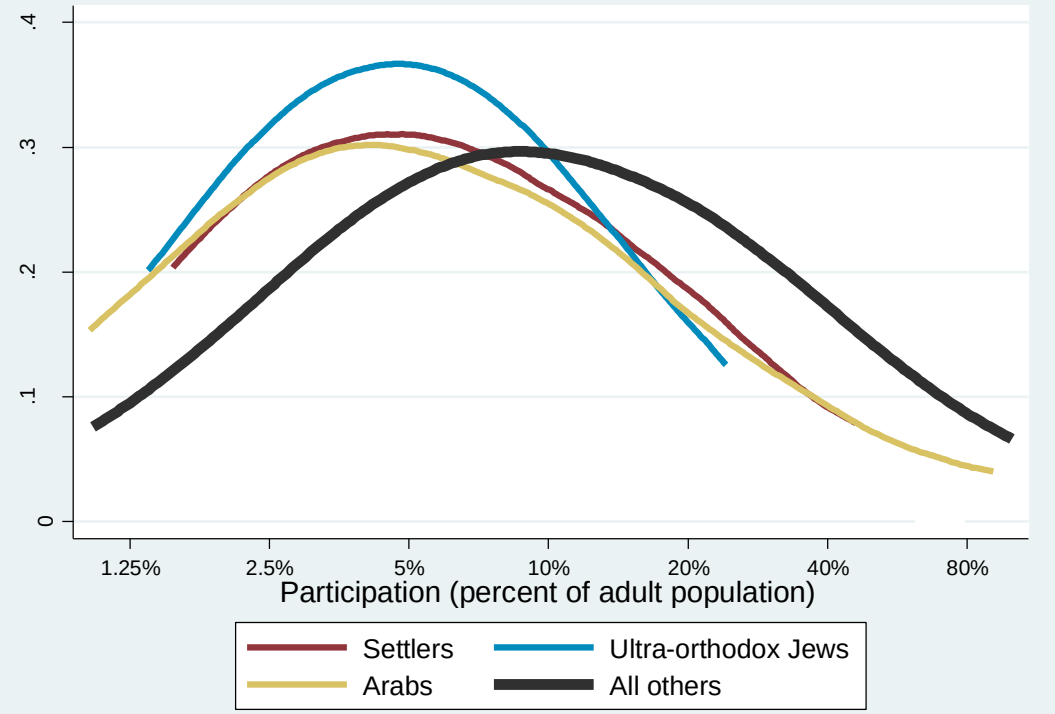
Map A6.3 does not mark the locations of small events not included in the dataset. The large bubbles representing the four epicenters are proportional to the size of the demonstrations they hosted during the protest. All remaining bubbles are uniform in size.

Appendix 4

The distribution of participation rates

Chart A4.1

Kernel density plot of participation rates by population sector



Notes: "Active" home localities only (peak participation rate of at least 1%). X-axis uses a log scale (base 2).

Appendix 5

Comparing behavioral and survey data

In general, where comparisons can be made, the findings presented in the paper are consistent with the results of public opinion polls conducted among representative samples of the adult population.[3] Yet we argue in Part 1 of the paper that estimates of participation obtained from surveys suffer from inadequate coverage of hard-to-reach populations. In order to test this claim, we compare participation rates of the three alienated sectors based on our behavioral data with parallel estimates based on two independent population surveys. These are the 2011 Taub Center Social Survey, carried out in mid-September 2011, and the Israel Democracy Institute's 2012 "Democracy Index" survey (https://en.idi.org.il/publications/8769) fielded in April and May 2012.[4]

In light of the methodological differences between surveys and our data, including their different units of analysis, we compare intergroup differentials rather than group-specific participation rates. These differentials are measured by odds ratios for each sector compared to the rest of the population. We followed this procedure not only for the minorities but also for a sub-population which we have no reason to suspect is misrepresented by surveys: the satellite cities of Tel Aviv, known collectively as the *Dan* district,[5] which have sociopolitical profiles indiciative of average or above-average

---

[3] See, for example, Hermann, Tamar, et al. 2012. *The Israeli Democracy Index 2012*. Jerusalem: The Israel Democracy Institute.

[4] The surveys have similar sample sizes that can be illustrated by the Taub Social Survey. Of the total of 1,002 adult respondents, 143 were Arabs, 85 defined themselves as Ultra-orthodox Jews, 32 lived in settlements and 204 resided in the Dan district. The following questions were posed: Taub Question 24 (Yes/No, trans. From Hebrew): "In July this year a wave of protests broke out demanding change in Israel's social policy. Did you participate in one or more of the demonstrations?". Democracy Index Qu. 75 (Yes/No, trans. From Hebrew): "Did you participate or not in one or more of the protest events last summer?".

[5] The 14 satellite cities in the Dan district are large enough that many of them were sampled in both surveys.

engagement in the protest. Mobilization of the residents of this disrict was contrasted with all other areas of the country, setting aside minority individuals and localities.[6]

The results of this analysis, presented in Chart A5.1, show a consistent pattern for the three minority sectors. In five out of six possible comparisons, the odds ratios based on our location data are much lower than those yielded by the two surveys. Given the large samples and automatic generation of the behavioral data, it is reasonable to conclude that the protest participation of minorities was overestimated in the polls, especially among settlers. Moreover, differences between the results for the two surveys (particularly striking for the Ultra-orthodox sector), and the implausible finding of both surveys that the participation of settlers was similar to the majority, reinforce our doubts about the ability of population surveys to reliably sample minority groups. It bears emphasis that since these three sectors live exclusively or predominantly in segregated localities, our locality-based behavioral estimates are free from suspicion of erroneous ecological inference.

A further weakness of population surveys is that without over-sampling for minorities, estimates suffer from marked uncertainty. The Confidence Intervals shown in the chart are extremely wide, yet despite that they do not always overlap with the results obtained using behavioral data (a failing that is especially noticeable for settlers).

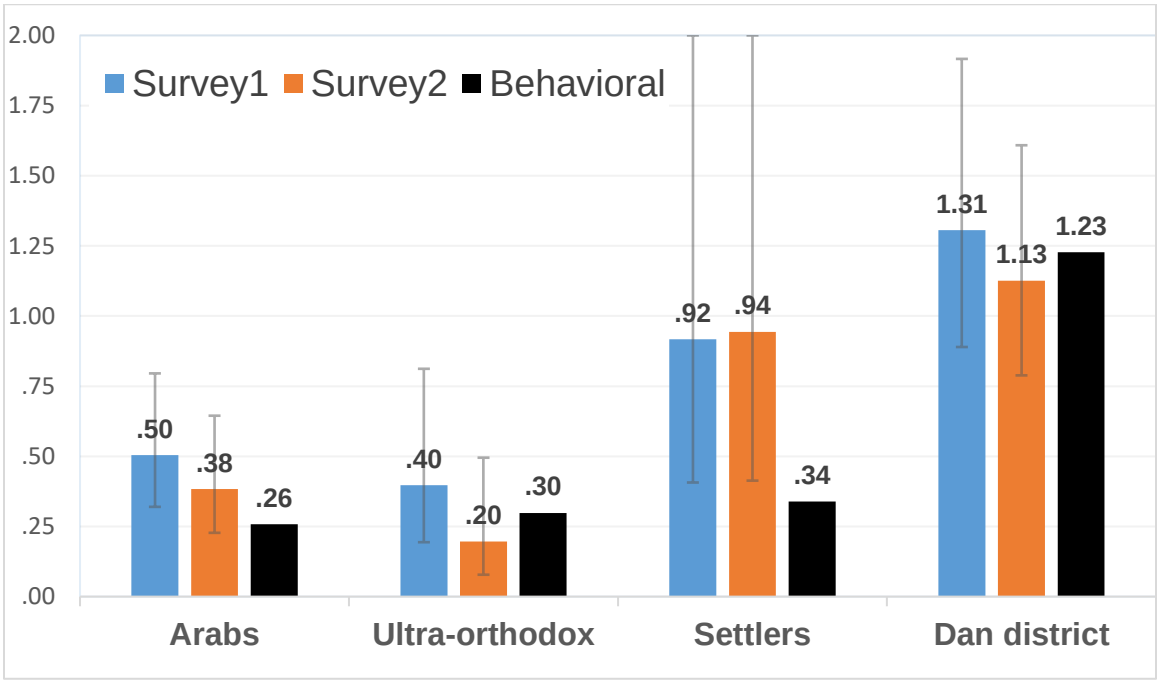
As expected, results for the Dan district confirm that differences between behavioral and survey data can be negligible when looking at a population that is unlikely to be misrepresented in surveys. It appears that national sample surveys may accurately report differences in protest engagement between large and well-sampled social sectors, but are limited in their ability to reach minority groups. One reason is that in cases like Israel, where many of the leaders of the minority communities were critical of

[6] Note that Tel Aviv itself, the major epicenter of the protest, was excluded from this part of the analysis because our behavioral estimates for events in Tel Aviv were more vulnerable to error in distinguishing protesters from bystanders.

the protest campaign, recruiting representative samples and obtaining frank answers is particularly difficult. These obstacles are avoided when protest behavior is measured directly and *in situ*.

Chart A5.1

Likelihood of participation based on surveys and behavioral data (Odds Ratios)



Notes: The results for surveys are based on data for individuals while those using behavioral location data are based on home locality participation rates. For any given subgroup, their relative likelihood of participation is the ratio of their odds of participating versus those of the majority (excluding Arab, Ultra-orthodox and settler individuals or localities). 95% Confidence Intervals are shown for individual-level odds ratios. For further details, see the text of this appendix.

Appendix 6

Multivariate analysis results[a]

Active/Inactive (n=2,172)	b		S.E.
Minorities	-.550	**	(.215)
Distance	-.067	**	(.006)
Minorities*Distance	-.005		(.018)
%left/center	.004		(.003)
%college	.023	**	(.005)
Constant	.406	**	(.139)
Peak participation rate (n=1,317)	b		S.E.
%left/center (power .5)	-1.46	**	(.22)
%left/center	.55	**	(.06)
%college	.01	*	(.00)
Distance (power -1)	4.54E-07	**	(.00)
Arabs	-.49	**	(.18)
Ultra-orthodox	-.95	**	(.19)
Settlers	-.23		(.17)
Constant	-2.67		(.04)
Left/center vote shift (n=1,300)	b		S.E.
%participation	7.88	**	(.63)
%college	2.72	**	(.29)
%college (squared)	-.31	**	(.05)
Constant	4.60	**	(.14)

Yesh Atid vote (n=1,301)	b		S.E.
%participation	21.21	**	(4.31)
%participation (power .5)	-22.18	**	(4.43)
%college	7.32	**	(.33)
%college (power 3)	-.10	**	(.01)
Constant	18.37	**	(.32)
Protest participation	b		S.E.
Individual level (n=659)			
Left/center	-.18		(.31)
Age	-.03	**	(.01)
BA	.69	**	(.20)
Religiosity [b]	1.27	**	(.32)
FSU immigrant	-1.50	**	(.46)
Female	-.38	*	(.19)
Constant	-1.19	**	(.38)
Locality level (n=118)			
Local mobilization	1.18		(2.17)
Local mobilization*Left/center	8.65	**	(3.02)

[a] See text for description of model specifications

[b] Secular & Traditional vs. Orthodox & Ultra-orthodox

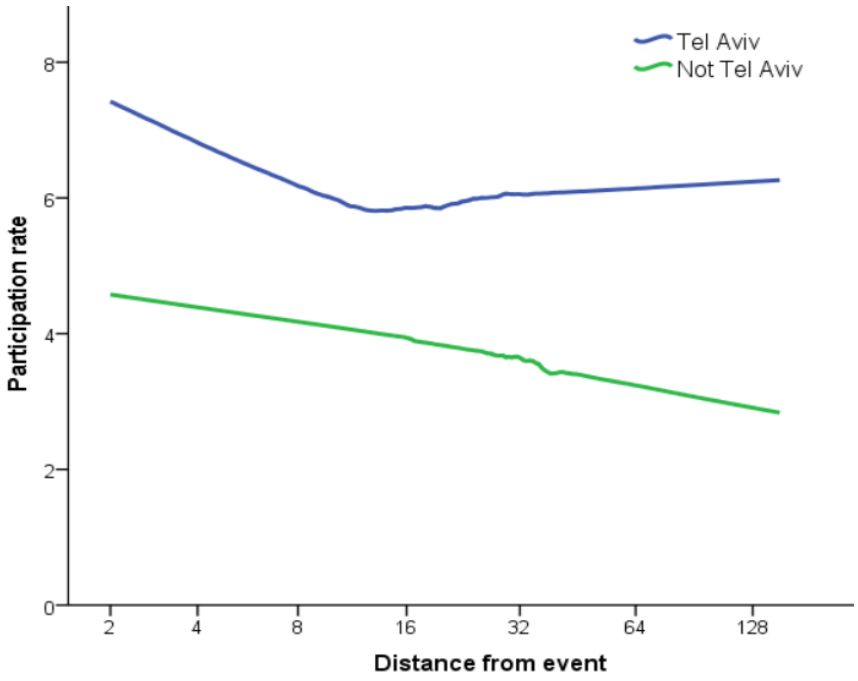
* p<0.05; ** p<0.01

Appendix 7

Distance effects for Tel Aviv and all other host locations

Chart A7.1

Effect of home localities' distance from protest events on their total participation, by event location (Tel Aviv vs. all other)



Notes: Loess-smoothed total participations of all home localities at all events, relative to their working-age population. Limited to participation rates of at least 1% in events held outside the home locality (n=2,380 combinations of home localities and events). X-axis shows distance in kilometers on a log scale (base 2), after bottom-coding at 2km and top-coding at 150km.