When does Descriptive Representation become Substantive?

Systematic Luck, Social Power and Resources' Allocation

Robustness Checks

In the main paper we use the statistical areas' socio-economic cluster rank as my measurement level which means that in the dataset we are re-counting the same observations measured on a city level. This could potentially inflate measurements, have an artificial effect on statistical variance and might eventually yield biased estimates. In this supplementary file we offer robustness checks for the results using alternative structuring of the data and then re-estimate the data using the same methods we use in the main paper. we basically run three series of checks: we first use the city's socio-economic rank as my nesting/hierarchical/random effect variable, we then use the city per-se as a nesting/hierarchical/random effect variable. Finally, we go back to a statistical areas' dataset and use the mixed cities' statistical areas' 496 categories socioeconomic rank (and not the cluster rank) as a nesting/hierarchical/random effect variable. All in all, the tests show that the main paper's results still hold. Yet, using the cities' socio-economic rank for nesting the models yields a too high to ignore levels of variance which are unaccounted for by this measurement level. Using the city as a nesting variable does not allow for valid theoretical inferences. Finally using the areas' socio-economic rank improves the model estimation yet offers outcomes which are not stable enough to process. Thus, the tests here validate my decisions to use the statistical areas as my unit of analysis, and to use the 22 categories areas' socio-economic cluster rank as the nesting variable.

Cities' Socio-Economic Rank as a 'Random' Effect

In order to test the data and hypotheses using this nesting variable we changed the dataset to be a city-year based data. Thus, each unit of analysis is a city in a given year, with each variable's either relate to the locality observation in a given year, or (when possible) to the statistical areas' observations annual. For that purpose, we re-shaped the dataset from the statistical area nested in a city level to a year-locality level and averaged all the statistical areas data for each year. The statistical procedures we use here are the same that we use in the main paper for the statistical area data. Please see the main paper's data and variables section for the variables' measurement and operational definitions.

Dependent Variable: New Houses in Stat. Areas	Model 1: Representation	Model 2: Ethnicity and Economics	Model 3: Politics
(Intercept)	7.56	30.99**	0.01
	(4.34)	(9.89)	(8.33)
Arab Reps %	56.03		
	(31.68)		
Arabs Ratio City		-0.67	
		(0.46)	
City Fiscal Strength		-2.56	
		(1.33)	
Turnout %			0.25
			(0.13)
Competition Council			-1.43
			(0.79)
Competition Mayor			0.52***
			(0.12)
Model Parameters			
AIC	1211.90	1100.32	1172.56
BIC	1223.40	1114.21	1192.58
Log Likelihood	-601.95	-545.16	-579.28
Num. Obs.	131	119	129
Num. Groups: Rank.Locality	29	29	29
Var: Rank.Locality (Intercept)	101.47	147.75	83.61

Var: Residual	500.82	460.50	405.35
Ran. Eff. Intercept	10.07	12.16	9.144
Ran. Eff. Residual	22.38	21.46	20.333

Table 1: City Year Measures for Housing

As in the models in the main paper (and in fact in a much clearer manner) two outcomes become apparent:

1. More Arab representatives in the council mean more new houses in the city. That said here this result yields a standard deviation which is too large to be significant.
2. The city's fiscal strength and the ratio of Arabs in it decrease new housing in the city.
3. The mayor's political power is highly effective in explaining more houses in the city.

Dependent Variable: New Infrastructure in Cities	Model 1: Representation	Model 2: Ethnicity and Economics	Model 3: Politics
(Intercept)	6.96***	10.62***	4.51
	(1.63)	(2.93)	(2.58)
Arab Reps %	9.72		
	(10.07)		
Arabs Ratio City		-0.27	
		(0.14)	
City Fiscal Strength		0.44	
		(0.38)	
Turnout %			-0.09*
			(0.04)
Competition Council			0.52***
			(0.13)
Competition Mayor			0.03
			(0.04)
Model Parameters			
AIC	881.25	794.20	860.58
BIC	892.75	808.10	877.74
Log Likelihood	-436.63	-392.10	-424.29
Num. Obs.	131	119	129
Num. Groups: Rank.Locality	29	29	29
Var: Rank.Locality (Intercept)	33.43	19.80	14.70

Var: Residual	32.03	31.95	33.51
Ran. Eff. Intercept	5.782	4.449	3.384
Ran. Eff. Residual	5.659	5.652	5.789

Table 2: City Year Measures for Infrastructure

As in the main paper, the main results (which in this case remain significant) stem from the political model showing that a decrease in competition in council increases the amount of new infrastructure in the cities. The Arab representatives' result flips yet is too varied to be significant. The same goes for the ethnicity and economics model, where also the standard deviation increases and the direction of influence of cities' self-funding flips.

Dependent Variable: Arabs' Annual Bagrut Eligibility	Model 1: Representation	Model 2: Ethnicity and Economics	Model 3: Politics
(Intercept)	49.15***	66.61***	40.95***
	(2.38)	(4.32)	(4.12)
Arab Reps %	48.11**		
	(15.45)		
Arabs Ratio City		-0.42*	
		(0.20)	
City Fiscal Strength		-2.23***	
		(0.58)	
Turnout %			0.23**
			(0.07)
Competition Council			-1.00*
			(0.39)
Competition Mayor			0.03
			(0.06)
Model Parameters			
AIC	997.69	903.78	974.93
BIC	1009.19	917.68	994.95
Log Likelihood	-494.84	-446.89	-480.47
Num. Obs.	131	119	129
Num. Groups: Rank.Locality	29	29	29
Var: Rank.Locality (Intercept)	62.18	27.87	39.54
Var: Residual	82.22	88.51	78.51
Ran. Eff. Intercept	7.885	5.279	6.288
Ran. Eff. Residual	9.068	9.408	8.860

Table 3: City Year Measures for Education

In this table also the results are basically the same as in the main paper. The results here show smaller power yet besides the mayor's power in council all other results are still strong and significant. This is the exact same pattern of outcomes again even strengthening the main paper's outcomes: more Arab representatives, less city income more Bagrut eligibility for Arab students in the city.

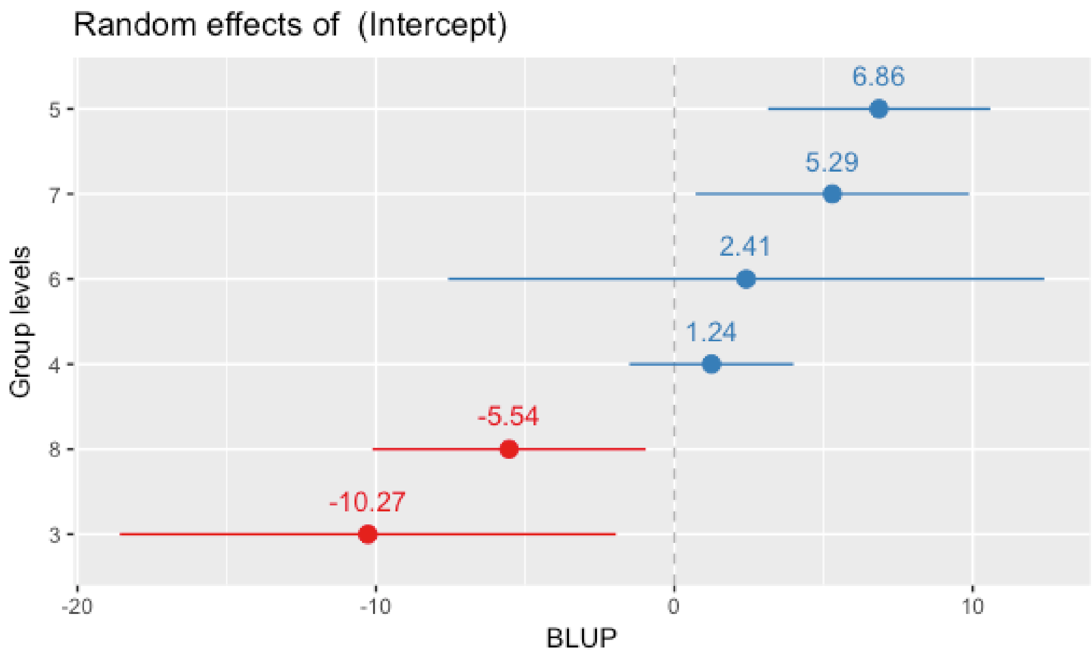

Figure 1: Representation and Education with Locality Cluster as Random Effect

Figure 1 shows what happens in this dataset or rather when one attempts to measure such a dynamic process as we do here while sticking to the city level rather than the statistical area level. Although all variables in this particular model are indeed city level the data shows that the representation effect is too diverse to be able to point at a concrete and significant direction. That is, when put in context of the cities' socio-economic rank (as this model does) the outcome is too diverse to be able to extract a clear outcome. Sub-terrain dynamic processes clearly take place within years and localities creating subtle changes that a locality socioeconomic rank-year design simply misses.

The City as the Random Effect

Another angle we examined in order to make sure that no stone is not unturned in making sure that we covered all alternative measurement and estimation options, is examining the variables using the city as the 'random' effect and see if we can find some type of explanation which is associated with the city per-se. An embedded theoretical assumption for that particular estimation is that there could be a city intrinsic factor that creates a set of outcomes which is a result of some city level characteristics such as size or location which outweighs social, economic and political factors in estimating the way resources are allocated to Arabs in mixed cities. **W**hen running the random effects associated with the cities one gets the following results:

City	Intercept
Ako	-4.464
Haifa	6.050
Lod	-11.785
Maalot-Tarshiha	2.217
Nazareth-Illit	4.008
Ramla	-12.374
Tel-Aviv Jaffa	16.348

Table 5: Random effects for Cities as Intercepts

Visualizing these results using again the sjplot R package we get the following hierarchy of intercept effects on the way representation affects education levels.

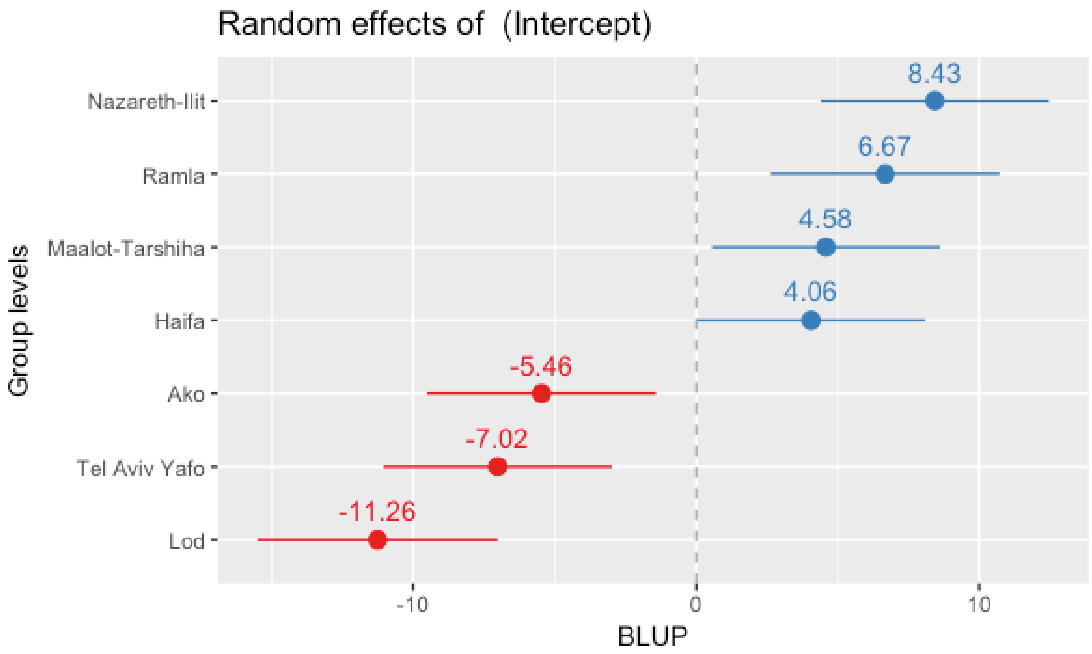

Figure 2: Representation and Education with Locality Name as a Random Effect

These results show that cities as different geographically and demographically as Ako and Tel-Aviv Yafo have a negative effect on the connection between the Arab representatives' ratio on city councils and the Bagrut eligibility rates. While also different cities such as Ramla and Haifa have a positive effect on that connection. Clearly, cities per-se have an effect on the representation effect strength and point of departure. Yet, that effect is unclear until one associates a city's intercept with its socio-economic rank. However, as the previous analysis showed when resorting to the socio-economic rank using the city-year level measures then the results become too vague and diverse to interpret.

Thus, the strategy we use in the paper of repeated measures of city-year observations on the basis of the cities' statistical areas, provides an estimation process that is both theoretically based and accounts for the variance in the data thereby allowing a proper estimation process for it. We now turn to illustrate the same tests using this time the areas' socio-economic rank as defined by the ICBS rather than using the ICBS' areas' socio-economic cluster rank.

Statistical Areas' Socio-Economic Rank as a 'Random' Effect

Dependent Variable: Housing	Model 1: Representation	Model 2: Ethnicity and Economics	Model 3: Politics
(Intercept)	8.33***	10.74**	3.36
	(1.42)	(3.80)	(3.53)
Arab Reps %	24.54		
	(14.45)		
Arabs Ratio City		0.18	
		(0.17)	
Arabs Ratio Area		-0.11**	
		(0.04)	
City Fiscal Strength		-0.19	
		(0.50)	
Turnout %			0.25***
			(0.06)
Competition Council			-0.48*
			(0.19)
Competition Mayor			0.14***
			(0.03)
Parties in Council			0.18
			(0.40)
Model Parameters			
AIC	64051.42	51965.19	64818.71
BIC	64078.28	52004.18	64865.81

Log Likelihood	-32021.71	-25976.59	-32402.36
Num. Obs.	6096	4914	6180
Num. Groups: Area Rank	496	496	496
Var: Area Rank (Intercept)	171.71	190.17	181.09
Var: Residual	2025.91	2150.80	1980.67

Table 6: New Housing with Stat. Areas' Rank as a Nesting Variable

This first of three tables, to a large extant shows findings similar to those revealed by table 4 in the main paper. The fit measures and significance outcomes seem to be even more efficient than the models in the main paper.

Dependent Variable: Infrastructure City	Model 1: Representation	Model 2: Ethnicity and Economics	Model 3: Politics
(Intercept)	13.98***	21.57***	5.76***
	(0.26)	(0.61)	(0.63)
Arab Reps %	-29.06***		
	(2.60)		
Arabs Ratio City		-0.58***	
		(0.03)	
Arabs Ratio Area		-0.01	
		(0.01)	
City Self-Funding Ratio		-0.99***	
		(0.08)	

Turnout %			0.14***
			(0.01)
Competition Council			0.66***
			(0.03)
Competition Mayor			-0.14***
			(0.01)
Parties in Council			-0.73***
			(0.07)
Model Parameters			
AIC	42934.13	34321.46	42693.44
BIC	42960.99	34360.46	42740.54
Log Likelihood	-21463.06	-17154.73	-21339.72
Num. Obs.	6096	4914	6180
Num. Groups: Area Rank	496	496	496
Var: Area Rank (Intercept)	5.92	2.84	11.48
Var: Residual	63.16	60.76	52.99

Table 7: New Infrastructures with Stat. Areas' Rank as a Nesting Variable

The same outcomes pattern happens when estimating new infrastructure in the city. Comparing results to table 5 in the main paper yields the same results and better fit models than those displayed in the main paper.

Dependent Variable: Arabs' Annual Bagrut Eligibility	Model 1: Representation	Model 2: Ethnicity and Economics	Model 3: Politics
(Intercept)	49.26***	72.36***	32.24***
	(0.45)	(0.82)	(0.83)
Arab Reps %	22.48***		
	(4.14)		
Arabs Ratio City		-0.75***	
		(0.04)	
Arabs Ratio Area		0.00	
		(0.01)	
City Self-Funding Ratio		-3.21***	
		(0.11)	
Turnout %			0.19***
			(0.01)
Competition Council			-0.03
			(0.05)
Competition Mayor			-0.07***
			(0.01)
Parties in Council			1.20***

			(0.10)
	Model Parameters		
AIC	45769.49	36154.79	45690.64
BIC	45796.35	36193.79	45737.74
Log Likelihood	-22880.75	-18071.39	- 22838.32
Num. Obs.	6096	4914	6180
Num. Groups: Area Rank	496	496	496
Var: Area Rank (Intercept)	37.08	17.74	26.90
Var: Residual	93.20	81.99	84.14

Table 8: Bagrut Eligibility with Stat. Areas' Rank as a Nesting Variable

Also the same pattern of mostly better estimates and more fit models happens also here when estimating the models using the statistical area's socio-economic rank rather than its cluster ranks. Obvious question is then: why not use this estimate as the paper's 'nesting' variable? To offer an answer please view figure 3 below.

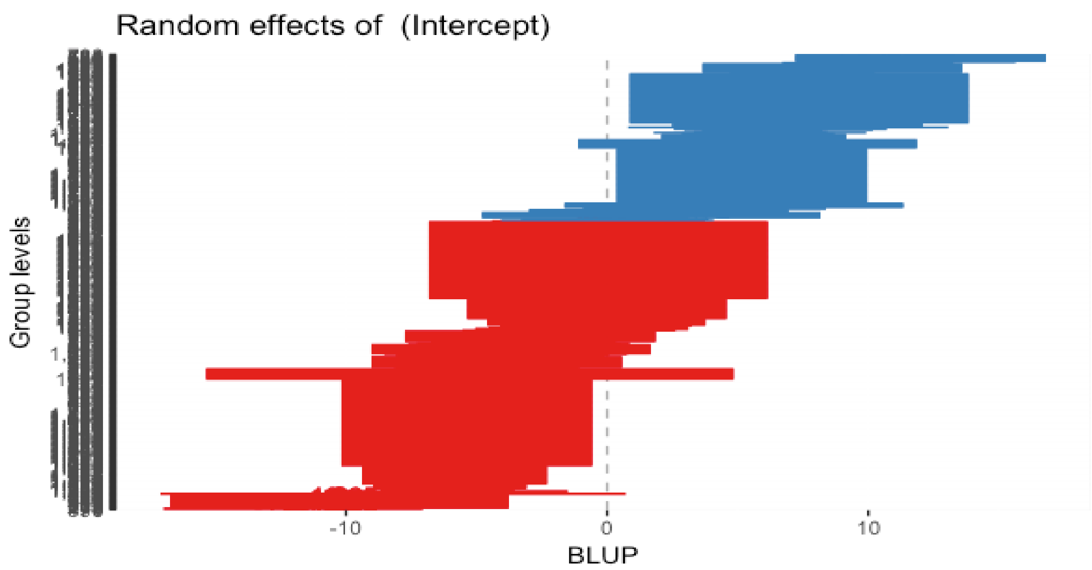

Figure 3: Random Effects Intercepts for Areas on Representation and Education

We use again the sjplot package for visualizing the random effects associated with the models we estimated above. The effects with the diversity of almost 500 groups of area ranks become too complex to draw a theoretically valid inference which could have some clarity in estimating the effects of social contexts on behaviors.

In conclusion and following all the tests we did here, we prefer using the more parsimonious measure of socio-economic clusters that I use in the main paper which as we see it includes the best of all worlds for this analysis: it is able to grasp social complexity and dynamics yet is parsimonious enough for drawing theoretical inferences from the statistical analysis.