

Appendix

Contents

T1	Sequential build-up of model	2
T2	Predicted probabilities for positive evaluation and moderators	8
T3	Robustness of “lookback” window for “About-to-be candidate”	9
T4	Fixed effects and two step evaluation process	13
T5	Matching	15
T6	Timing of evaluation decision.....	23
T7	Exclusion of alternative explanations	25
T8	Description of post-hoc analysis	29
T9	Test of multinomial categories	31
T10	Overview of the RfA process	32

T1 Sequential build-up of model

This section shows the sequential build-up of our main regression model. We also include the results for the neutral category.

We start by including only our main variables (independent and moderator) and then add control variables on the level of the evaluating member, evaluated peer, dyad, the RfA process, and finally, year fixed effects. The results are shown in Table T1 and T2.

Our main results remain relatively constant. Only the interaction of About-to-be candidate and Candidate Activity is strongly influenced by the control variables. Looking at the change of R-squared, we find the evaluating member, evaluated peer, and dyad add roughly the same amount of explanatory power. By far, the strongest increase in R-squared is by accounting for the dynamics of the RfA process. One interpretation is that members pay much more attention to how others voted than to the characteristics of the candidate (or other factors).

Table T1: Multinomial logistic regression; DV: decision how to evaluate (base case: abstention), with neutral category, part 1

	Neutral	Negative	Positive	Neutral	Negative	Positive	Neutral	Negative	Positive
About-to-be candidate (ATBC)	1.435*** (0.141)	1.389*** (0.094)	1.216*** (0.080)	1.512*** (0.121)	1.603*** (0.108)	1.179*** (0.092)	1.459*** (0.119)	1.474*** (0.109)	1.204*** (0.092)
Evaluation Pivotality	0.181 (0.118)	0.534** (0.040)	-0.968*** (0.031)	0.212** (0.067)	0.427*** (0.040)	-0.952*** (0.033)	0.313*** (0.056)	0.604*** (0.038)	-0.946*** (0.032)
Candidate Activity	-1.318*** (0.201)	-1.242*** (0.108)	-0.104 (0.071)	-1.551*** (0.149)	-1.598*** (0.104)	-0.419*** (0.070)	-0.326 (0.218)	-0.660*** (0.156)	-0.348** (0.110)
ATBC * Evaluation Pivotality	-0.174 (0.145)	-0.690*** (0.088)	-0.229** (0.071)	-0.227* (0.114)	-0.754*** (0.097)	-0.184* (0.077)	-0.253* (0.108)	-0.714*** (0.096)	-0.174* (0.078)
ATBC * Candidate Activity	-1.416** (0.489)	-2.947*** (0.397)	-0.159 (0.218)	-1.490** (0.486)	-2.867*** (0.411)	-0.008 (0.243)	-0.859* (0.417)	-1.569*** (0.342)	-0.028 (0.258)
Evaluating member:									
Activity				1.007*** (0.296)	0.411 (0.222)	0.290 (0.195)	1.048*** (0.307)	0.480* (0.229)	0.423* (0.196)
Edited Pages				-0.059 (0.083)	-0.156** (0.059)	-0.144** (0.055)	-0.096 (0.083)	-0.188** (0.059)	-0.171** (0.055)
Edits				0.018 (0.080)	0.068 (0.056)	0.049 (0.053)	0.069 (0.080)	0.124* (0.057)	0.023 (0.053)
Engagement				-0.074 (0.039)	0.085*** (0.022)	-0.111*** (0.027)	-0.127** (0.043)	0.015 (0.028)	0.016 (0.021)
Share of Edits				0.701* (0.309)	-0.050 (0.243)	0.694** (0.259)	0.704* (0.303)	-0.012 (0.243)	0.650* (0.266)
Tenure				0.005** (0.002)	0.007*** (0.001)	0.005*** (0.001)	0.008*** (0.002)	0.012*** (0.001)	0.005*** (0.001)
No. of Elections Participated				-0.001*** (0.000)	-0.001 (0.000)	-0.001 (0.000)	-0.002*** (0.000)	-0.001* (0.000)	-0.001* (0.000)
Admin				-0.026 (0.077)	0.097 (0.062)	-0.051 (0.053)	-0.024 (0.077)	0.088 (0.063)	-0.049 (0.053)
Relative Time to Vote				0.340* (0.171)	0.418** (0.133)	0.620*** (0.119)	0.433* (0.170)	0.525*** (0.135)	0.616*** (0.118)
Previous Failed Candidacy				0.176 (0.127)	0.181 (0.110)	0.208* (0.097)	0.156 (0.127)	0.178 (0.111)	0.191 (0.099)
Reversed Evaluations				0.004*** (0.000)	0.001 (0.000)	0.001*** (0.000)	0.004*** (0.000)	0.000 (0.000)	0.002*** (0.000)
Deviation from Majority				0.230 (0.210)	3.651*** (0.162)	-1.703*** (0.155)	0.235 (0.209)	3.859*** (0.164)	-1.813*** (0.152)
Evaluated peer:									
Edited Pages (Log.)							-0.081* (0.036)	-0.216*** (0.030)	0.138*** (0.021)
Edits (Log.)							0.428*** (0.058)	0.438*** (0.041)	-0.077* (0.036)
Tendency to Evaluate Neg.							1.456* (0.643)	3.866*** (0.456)	0.782* (0.324)

Engagement							-0.029 (0.020)	-0.025 (0.018)	-0.058*** (0.013)
Share of Edits							-1.205*** (0.215)	-1.466*** (0.163)	-0.678*** (0.137)
Tenure							-0.011*** (0.001)	-0.012*** (0.001)	-0.004*** (0.001)
Edits Commented							-0.561*** (0.094)	-1.036*** (0.067)	0.725*** (0.070)
Edits Reverted (Total)							0.000** (0.000)	0.001*** (0.000)	-0.000 (0.000)
Edits Reverted (Share)							0.059 (0.145)	-0.349** (0.108)	-0.078 (0.083)
Self-nomination							0.465*** (0.035)	0.483*** (0.025)	-0.284*** (0.016)
Previous Failed Candidacy							0.261*** (0.028)	0.436*** (0.020)	0.056*** (0.015)
Edits, Content pages							-0.171*** (0.028)	-0.208*** (0.020)	0.129*** (0.018)
Edits, Discussion pages							-0.032** (0.011)	-0.037*** (0.009)	0.014* (0.007)
Edits, Template pages							0.001 (0.009)	0.025*** (0.006)	-0.005 (0.004)
Edits, Categories pages							-0.032*** (0.009)	-0.043*** (0.007)	-0.012* (0.005)
Edits, User pages							0.020 (0.010)	0.023** (0.008)	-0.023*** (0.006)
Edits, User discussion							0.001 (0.019)	0.038** (0.014)	0.027* (0.011)
Edits, Wikipedia pages							-0.132*** (0.018)	-0.152*** (0.013)	0.101*** (0.010)
Edits, Files pages							0.006 (0.008)	-0.002 (0.005)	-0.004 (0.004)
Activity Gaps							0.002*** (0.000)	0.003*** (0.000)	-0.002*** (0.000)
Created Pages (Log.)							-0.017 (0.027)	0.002 (0.021)	-0.084*** (0.016)
Average Similarity							-81.444*** (5.566)	-85.228*** (4.071)	-57.901*** (3.724)
Constant	-2.867*** (0.108)	-1.816*** (0.036)	-0.184*** (0.027)	-3.013*** (0.240)	-2.235*** (0.151)	0.439** (0.141)	-2.605*** (0.339)	-0.703*** (0.208)	-0.529** (0.202)
Pseudo R-squared		0.029			0.062			0.096	
Log-likelihood		-378,757			-346,838			-330,389	
N		298,921			277,232			274,674	

Note: Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table T2: Multinomial logistic regression; DV: decision how to evaluate (base case: abstention), with neutral category, part 2

	Neutral	Negative	Positive	Neutral	Negative	Positive	Neutral	Negative	Positive
About-to-be candidate (ATBC)	1.470*** (0.119)	1.482*** (0.110)	1.223*** (0.093)	1.315*** (0.113)	1.269*** (0.103)	1.128*** (0.096)	1.350*** (0.113)	1.305*** (0.102)	1.154*** (0.094)
Evaluation Pivotality	0.314*** (0.056)	0.600*** (0.038)	-0.933*** (0.033)	1.299*** (0.064)	1.673*** (0.049)	0.600*** (0.043)	1.316*** (0.063)	1.683*** (0.050)	0.614*** (0.043)
Candidate Activity	-0.673** (0.217)	-0.812*** (0.157)	-0.973*** (0.118)	-0.690** (0.236)	-0.736*** (0.199)	-0.604*** (0.155)	-0.888*** (0.242)	-0.902*** (0.194)	-0.847*** (0.150)
ATBC * Evaluation Pivotality	-0.294** (0.108)	-0.743*** (0.097)	-0.232** (0.079)	-0.295** (0.113)	-0.611*** (0.102)	-0.384*** (0.089)	-0.285* (0.115)	-0.602*** (0.103)	-0.370*** (0.090)
ATBC * Candidate Activity	-0.938* (0.408)	-1.679*** (0.341)	-0.027 (0.258)	-0.324 (0.429)	-0.859* (0.369)	0.116 (0.316)	-0.340 (0.434)	-0.889* (0.373)	0.135 (0.319)
Evaluating member:									
Activity	1.111*** (0.308)	0.573* (0.232)	0.480* (0.200)	0.490 (0.286)	0.019 (0.239)	0.000 (0.208)	0.349 (0.293)	-0.104 (0.252)	-0.196 (0.217)
Edited Pages	-0.125 (0.084)	-0.243*** (0.060)	-0.166** (0.057)	-0.146 (0.083)	-0.254*** (0.063)	-0.177** (0.059)	-0.148 (0.082)	-0.254*** (0.063)	-0.173** (0.058)
Edits	0.085 (0.081)	0.175** (0.058)	-0.017 (0.055)	0.088 (0.083)	0.173** (0.063)	-0.028 (0.057)	0.079 (0.082)	0.163** (0.062)	-0.034 (0.057)
Engagement	-0.153*** (0.043)	-0.010 (0.028)	-0.025 (0.023)	-0.177*** (0.046)	-0.012 (0.027)	-0.063** (0.023)	-0.147** (0.049)	0.014 (0.028)	-0.045 (0.024)
Share of Edits	0.513 (0.307)	-0.302 (0.244)	0.622* (0.268)	0.415 (0.311)	-0.386 (0.254)	0.707** (0.256)	0.313 (0.316)	-0.484 (0.254)	0.578* (0.263)
Tenure	0.008*** (0.002)	0.012*** (0.001)	0.005*** (0.001)	0.011** (0.002)	0.015*** (0.002)	0.010*** (0.002)	0.007** (0.002)	0.012*** (0.002)	0.006** (0.002)
No. of Elections Participated	-0.002*** (0.000)	-0.001* (0.000)	-0.001 (0.000)	-0.002*** (0.000)	-0.001* (0.000)	-0.000 (0.000)	-0.001*** (0.000)	-0.001 (0.000)	-0.000 (0.000)
Admin	-0.058 (0.078)	0.070 (0.063)	-0.110* (0.055)	-0.023 (0.079)	0.094 (0.066)	-0.088 (0.057)	0.008 (0.079)	0.122 (0.066)	-0.058 (0.057)
Relative Time to Vote	0.451** (0.170)	0.510*** (0.135)	0.687*** (0.121)	1.915*** (0.193)	1.830*** (0.153)	2.454*** (0.140)	1.964*** (0.189)	1.863*** (0.152)	2.495*** (0.138)
Previous Failed Candidacy	0.156 (0.127)	0.180 (0.111)	0.190 (0.100)	0.071 (0.123)	0.088 (0.110)	0.145 (0.106)	0.045 (0.122)	0.062 (0.109)	0.124 (0.103)
Reversed Evaluations	0.004*** (0.000)	0.000 (0.000)	0.002*** (0.000)	0.005*** (0.000)	0.001 (0.000)	0.002*** (0.000)	0.005*** (0.000)	0.001 (0.001)	0.002*** (0.000)
Deviation from Majority	0.212 (0.211)	3.837*** (0.165)	-1.835*** (0.154)	-0.335 (0.229)	3.831*** (0.170)	-2.117*** (0.155)	-0.359 (0.227)	3.834*** (0.169)	-2.128*** (0.155)
Evaluated peer:									
Edited Pages (Log.)	-0.096**	-0.230***	0.116***	-0.005	-0.070*	-0.063**	0.033	-0.034	-0.026

Edits (Log.)	(0.036) 0.446***	(0.030) 0.459***	(0.022) -0.065	(0.035) 0.380***	(0.029) 0.274***	(0.024) 0.296***	(0.035) 0.346***	(0.029) 0.248***	(0.023) 0.246***
Tendency to Evaluate Neg.	(0.058) 1.820**	(0.041) 4.002***	(0.037) 1.637***	(0.058) 2.713***	(0.042) 4.921***	(0.042) 2.394***	(0.062) 2.230**	(0.043) 4.458***	(0.040) 1.935***
Engagement	(0.650) -0.045*	(0.458) -0.039*	(0.334) -0.086***	(0.785) -0.205***	(0.704) -0.168**	(0.541) -0.219**	(0.780) -0.182***	(0.682) -0.153***	(0.524) -0.191***
Share of Edits	(0.021) -1.230**	(0.018) -1.483***	(0.014) -0.703***	(0.023) -1.584***	(0.020) -1.639***	(0.014) -1.402***	(0.023) -1.518***	(0.021) -1.524***	(0.016) -1.355***
Tenure	(0.216) -0.011***	(0.164) -0.011***	(0.137) -0.004***	(0.203) 0.003**	(0.151) 0.003***	(0.133) 0.008***	(0.190) -0.003*	(0.141) -0.002*	(0.119) 0.002*
Edits Commented	(0.001) -0.553***	(0.001) -1.033***	(0.001) 0.737***	(0.001) -0.723***	(0.001) -1.034***	(0.001) -0.430***	(0.001) -0.841***	(0.001) -1.150***	(0.001) -0.575***
Edits Reverted (Total)	(0.094) 0.000**	(0.067) 0.001**	(0.071) -0.000	(0.097) 0.001***	(0.070) 0.001***	(0.087) 0.001***	(0.098) 0.001***	(0.071) 0.001***	(0.085) 0.001***
Edits Reverted (Share)	(0.000) 0.163	(0.000) -0.275*	(0.000) 0.042	(0.000) -0.027	(0.000) -0.356**	(0.000) -0.054	(0.000) -0.176	(0.000) -0.493***	(0.000) -0.176*
Self-nomination	(0.143) 0.467***	(0.107) 0.485***	(0.081) -0.280***	(0.145) 0.107**	(0.108) 0.026	(0.088) -0.131***	(0.146) 0.079*	(0.104) 0.008	(0.083) -0.158***
Previous Failed Candidacy	(0.035) 0.244***	(0.025) 0.423***	(0.016) 0.033*	(0.033) 0.327***	(0.024) 0.448***	(0.017) 0.323***	(0.035) 0.359***	(0.024) 0.478***	(0.017) 0.362***
Edits, Content pages	(0.029) -0.186***	(0.020) -0.224***	(0.015) 0.119***	(0.030) -0.144***	(0.023) -0.136***	(0.018) -0.083***	(0.030) -0.127***	(0.022) -0.120***	(0.018) -0.066***
Edits, Discussion pages	(0.028) -0.033**	(0.020) -0.038***	(0.019) 0.015*	(0.029) 0.052***	(0.021) 0.052***	(0.021) 0.068***	(0.031) 0.052***	(0.021) 0.051***	(0.019) 0.069***
Edits, Template pages	(0.011) -0.002	(0.009) 0.022***	(0.007) -0.010*	(0.012) 0.003	(0.010) 0.027***	(0.008) 0.009	(0.012) -0.008	(0.010) 0.015*	(0.008) -0.004
Edits, Categories pages	(0.009) -0.034***	(0.006) -0.045***	(0.005) -0.014**	(0.010) -0.025**	(0.007) -0.038***	(0.006) -0.017**	(0.009) -0.023*	(0.007) -0.038***	(0.006) -0.013**
Edits, User pages	(0.009) 0.022*	(0.007) 0.024**	(0.005) -0.021***	(0.009) 0.015	(0.007) 0.016	(0.005) -0.002	(0.009) 0.008	(0.007) 0.007	(0.005) -0.008
Edits, User discussion	(0.010) -0.002	(0.008) 0.036*	(0.006) 0.023*	(0.010) 0.034	(0.008) 0.070***	(0.006) 0.077***	(0.010) -0.016	(0.008) 0.020	(0.006) 0.029*
Edits, Wikipedia pages	(0.019) -0.137***	(0.014) -0.156***	(0.011) 0.093***	(0.020) -0.036	(0.015) -0.028*	(0.013) 0.059***	(0.018) -0.025	(0.014) -0.015	(0.012) 0.069***
Edits, Files pages	(0.019) 0.004	(0.013) -0.004	(0.010) -0.006	(0.018) -0.013	(0.013) -0.021***	(0.010) -0.020***	(0.018) -0.013	(0.013) -0.023***	(0.010) -0.021***
Activity Gaps	(0.008) 0.002***	(0.006) 0.003***	(0.004) -0.003***	(0.009) 0.003***	(0.006) 0.004***	(0.005) 0.003***	(0.008) 0.003***	(0.006) 0.004***	(0.005) 0.003***
Created Pages (Log.)	(0.000) -0.011	(0.000) 0.007	(0.000) -0.074***	(0.000) -0.033	(0.000) -0.035	(0.000) -0.049**	(0.000) -0.032	(0.000) -0.037	(0.000) -0.047**

Average Similarity	(0.027) -69.115*** (5.942)	(0.021) -62.407*** (4.317)	(0.017) -60.662*** (3.907)	(0.028) -74.790*** (6.002)	(0.021) -66.178*** (4.544)	(0.017) -88.129*** (4.478)	(0.026) -60.491*** (5.689)	(0.021) -52.108*** (4.426)	(0.016) -75.205*** (4.028)
Dyad:									
Preceding Peer Approval	2.783*** (0.129)	2.436*** (0.124)	3.041*** (0.112)	2.803*** (0.156)	2.476*** (0.149)	3.030*** (0.134)	2.822*** (0.154)	2.492*** (0.148)	3.051*** (0.134)
Preceding Peer Criticism	2.714*** (0.279)	2.838*** (0.277)	2.229*** (0.258)	2.685*** (0.401)	2.811*** (0.407)	2.168*** (0.366)	2.670*** (0.399)	2.794*** (0.405)	2.155*** (0.364)
Work Overlap	-0.109*** (0.016)	-0.169*** (0.013)	-0.023* (0.010)	-0.114*** (0.016)	-0.174*** (0.014)	-0.031** (0.011)	-0.114*** (0.016)	-0.174*** (0.014)	-0.031** (0.011)
Edits of Candidate Deleted	0.291*** (0.049)	0.191*** (0.039)	0.509*** (0.030)	0.214*** (0.054)	0.097* (0.045)	0.468*** (0.034)	0.217*** (0.054)	0.102* (0.046)	0.470*** (0.034)
Edits by Candidate Deleted	0.294*** (0.065)	0.270*** (0.061)	0.390*** (0.063)	0.238*** (0.059)	0.206*** (0.056)	0.344*** (0.051)	0.230*** (0.061)	0.202*** (0.058)	0.332*** (0.054)
Edits of Candidate Deleted * Edits by Candidate Deleted	-0.052 (0.032)	0.012 (0.027)	-0.161*** (0.024)	-0.031 (0.034)	0.031 (0.028)	-0.149*** (0.024)	-0.032 (0.034)	0.029 (0.028)	-0.147*** (0.025)
Election-level controls:									
Activity First Day of Election				0.044*** (0.001)	0.044*** (0.001)	0.046*** (0.001)	0.045*** (0.001)	0.044*** (0.001)	0.047*** (0.001)
Current no. Evaluations				-0.059*** (0.002)	-0.056*** (0.001)	-0.059*** (0.001)	-0.060*** (0.002)	-0.058*** (0.001)	-0.060*** (0.001)
Current Approval Rating				0.428*** (0.080)	-0.398*** (0.076)	4.333*** (0.095)	0.615*** (0.083)	-0.212** (0.077)	4.530*** (0.088)
Current Number of Comments				-0.007*** (0.001)	-0.006*** (0.001)	-0.010*** (0.001)	-0.006*** (0.001)	-0.005*** (0.001)	-0.010*** (0.001)
No. Elections Same Month				-0.002 (0.001)	0.002 (0.001)	-0.002** (0.001)	0.004* (0.001)	0.006*** (0.001)	0.004*** (0.001)
Constant	-2.499*** (0.341)	-0.715*** (0.210)	-0.179 (0.204)	-3.555*** (0.358)	-1.747*** (0.225)	-3.076*** (0.290)	-3.391*** (0.347)	-1.653*** (0.237)	-2.857*** (0.284)
Year fixed effects	No	No	No	No	No	No	Yes	Yes	Yes
Pseudo R-squared	0.113			0.341			0.345		
Log-likelihood	-324,180			-240,835			-239,497		
N	274,674			274,672			274,672		

Note: Robust standard errors in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001.

T2 Predicted probabilities for positive evaluation and moderators

Below we show the predicted probabilities for positive evaluations when our main independent variable (“About-to-be candidate”) is moderated by pivotality of the evaluation (H2a) activity of the candidate (H3a). Figure T1 shows that about-to-be-candidates become less likely to evaluate a candidate positively the more pivotal the evaluation is. In contrast, other members do not change their tendency to evaluate positively depending on the pivotality. Figure T2 depicts that candidate activity does not affect a about-to-be-candidate’s tendency to evaluate positively.

Figure T1: Predicted probability for positive evaluation by members moderated by pivotality (H2).

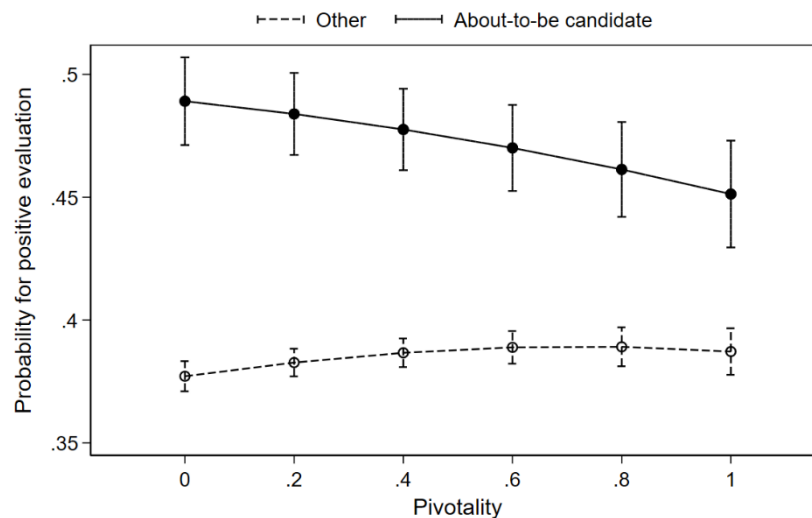
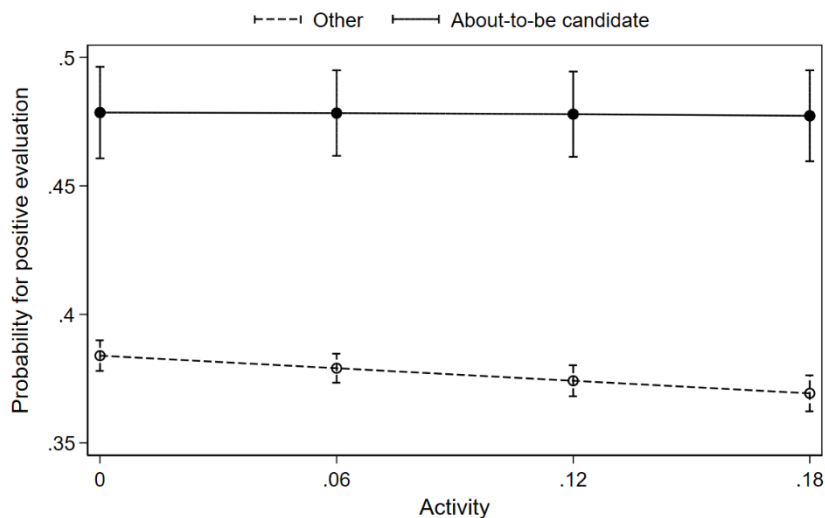


Figure T2: Predicted probability for positive evaluation by members moderated by candidate activity (H3).



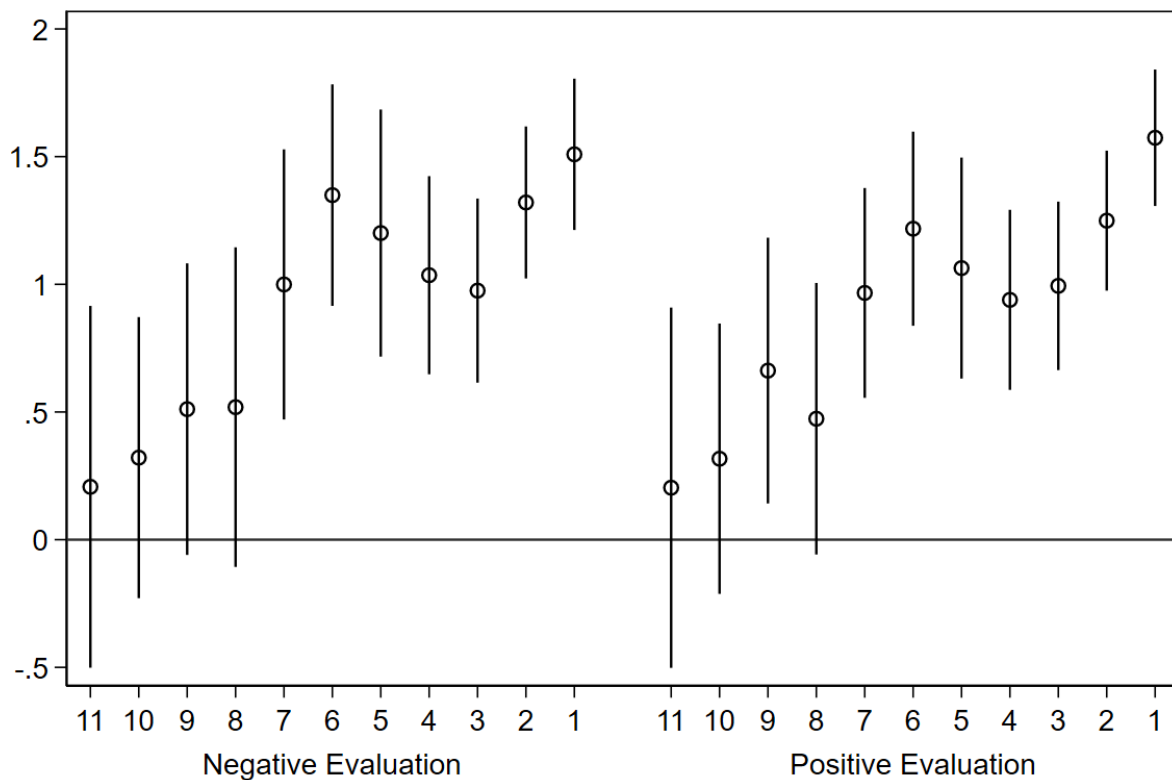
T3 Robustness of “lookback” window for “About-to-be candidate”

One assumption is that members *act (more) strategically* when approaching their own evaluation, and we estimate this falls in a window of about six months before their election. The timeframe was suggested in our interviews, as elaborated on further below. To determine whether this time frame is adequate, we perform two tests to assess whether this holds empirically.

T3.1 Anticipation of own evaluation

The first test rests upon the idea of creating dummies for all 11 months before the focal member’s RfA and running the main regression (Table 2) with the dummies instead of the “About-to-be candidate” variable. This approach allows us to see the months the effect differs from zero. The figure below draws all coefficients with 95 percent confidence intervals for affecting the likelihood to negatively and positively evaluate a candidate up to 11 months before the focal member’s RfA. For instance, for negative evaluations depicted on the left of the figure, we see that in months between month -1 and month -7 there is a significant effect. Our results indicate that the evaluating member is at least six months more likely to evaluate a candidate positively or negatively, that is, more active. In sum, these analyses and coefficients lend support to focusing on the six months prior.

Figure T3: Effect of “About to be candidate” in months leading to focal member’s RfA



T3.2 Time between next candidacy

Our second test to evaluate the time window uses the fact that some candidates have multiple candidacies. The idea is to see how long, on average, it takes someone to feel they can try again to become an administrator. The time window between candidacies indicates how long members believe they have to wait before a new and different assessment can be conducted, that is, how long members need before reappraising a candidate. Anecdotal evidence from comments during RfAs suggests that members should wait at least three or four months since a shorter window would not be sufficient to completely re-evaluate a candidate. We also compared the distribution (Figure T4) and summary statistics (Table T3) of the time (in days) between candidacies. The median is around 150 days (5 months), and the mean is around 240 days (8 months). Six months fall roughly into this window as well. Note that most candidates whose next candidacy is successful wait at least 90 days (3 months) before requesting adminship, which again points in the direction of using six months as the window.

Figure T4: Histogram of time between candidacies

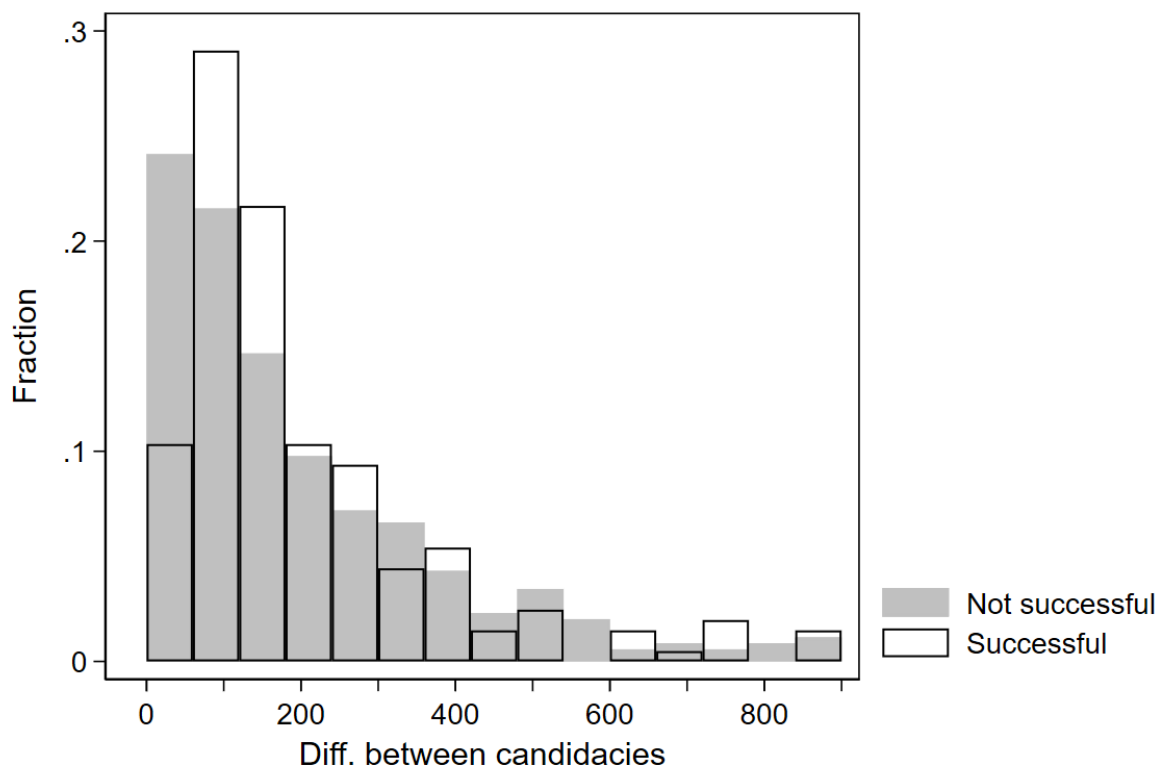


Table T3: Summary statistics about time between candidacies

	Mean	2 nd quartile	Median	4 th quartile	Std. dev.
Unsuccessful	261.6965	69	147	326	345.2179
Successful	241.1298	88	147.5	279	311.6105

T3.3 Three- and nine-month lookback window

The above two tests use features of the data to motivate the lookback window empirically. In addition to these precautions, we also test the robustness of the chosen six-month lookback window by running the main regression with the “About-to-be candidate” variable with a lookback window of three and nine months. The results remain qualitatively the same and quantitatively very similar (see Table T4). These tests show that our results are insensitive to whether we use a three-, six-, or nine-month lookback window.

Table T4: Multinomial logistic regression. DV: Decision on how to evaluate., 3,6, and, 9-month lookback window for “About-to-be candidate”

	Neutral	3 months Negative	Positive	Neutral	6 months Negative	Positive	Neutral	9 months Negative	Positive
About-to-be candidate (ATBC)	1.225*** (0.114)	1.213*** (0.107)	1.108*** (0.098)	1.350*** (0.113)	1.303*** (0.102)	1.153*** (0.094)	1.329*** (0.105)	1.255*** (0.093)	1.132*** (0.087)
Evaluation Pivotality	1.303*** (0.060)	1.671*** (0.049)	0.607*** (0.042)	1.315*** (0.063)	1.682*** (0.050)	0.615*** (0.043)	1.309*** (0.064)	1.682*** (0.049)	0.619*** (0.043)
Candidate Activity	-0.912*** (0.241)	-0.928*** (0.193)	-0.837*** (0.150)	-0.891*** (0.243)	-0.909*** (0.194)	-0.845*** (0.150)	-0.878*** (0.244)	-0.880*** (0.195)	-0.838*** (0.151)
ATBC * Evaluation Pivotality	-0.254* (0.127)	-0.683*** (0.116)	-0.392*** (0.108)	-0.285* (0.115)	-0.603*** (0.103)	-0.369*** (0.090)	-0.233* (0.109)	-0.536*** (0.097)	-0.350*** (0.090)
ATBC * Candidate Activity	-0.464 (0.520)	-1.238** (0.463)	-0.005 (0.399)	-0.343 (0.434)	-0.898* (0.374)	0.135 (0.319)	-0.407 (0.440)	-1.006** (0.373)	0.026 (0.309)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Candidate-level control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dyad-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Election-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared		0.343			0.364			0.364	
Log-likelihood		-240,118			-198,951			-198,888	
N		274,672			252,105			252,105	

Note: Robust standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

T4 Fixed effects and two-step evaluation process

A potential limitation of our empirical analysis could be that our multinomial regressions do not account for time-invariant unobservable evaluator fixed effects. Therefore, it is possible that evaluators do not become more active (positive and negative evaluations) due to anticipating their own evaluation, but more active evaluators want to become admins. In that case there is a selection effect that would explain strategic behavior.

Including fixed effects in multinomial regressions (and logistic regressions) is computationally still limited (e.g. the package *femlogit* for Stata only works with a limited set of fixed effect dummies). This is why we test the robustness of the results using a pair of linear regressions, i.e., linear probability models. First, we test the decision to participate and evaluate a candidate positively (instead of negatively). This separation also allows us to disentangle how much “About-to-be candidate” affects participation compared to the direction of the evaluation. (Table T5).

Anticipating an evaluation affects the tendency to evaluate positively and slightly shifts the evaluation from positive to negative. We also find that both moderators have a negative effect on participation and a positive effect on the type of evaluation (from negative to positive). Controlling for candidate fixed effects, we see an even stronger moderating effect of candidate activity.

Note that due to the two-step regression, we cannot determine whether changes between positive and negative evaluations are due to shifts from one to another or shifts from or to abstentions.

Table T5: Linear regression, DVs: Decision on whether or how to evaluate. Moderators included

	Evaluating member FE		Evaluated peer FE	
	Decision to vote (Voter)	Positive vs. Negative (Voter)	Decision to vote (Cand.)	Positive vs. Negative (Cand.)
About-to-be candidate (ATBC)	0.080*** (0.012)	-0.018** (0.006)	0.164*** (0.004)	-0.023*** (0.003)
Evaluation Pivotality	0.056*** (0.004)	-0.075*** (0.008)	0.088*** (0.017)	-0.096*** (0.009)
Candidate Activity	-0.064*** (0.013)	-0.081*** (0.017)	0.543* (0.236)	0.153 (0.138)
ATBC * Evaluation Pivotality	-0.040*** (0.008)	0.052*** (0.014)	-0.037*** (0.009)	0.046*** (0.009)
ATBC * Candidate Activity	-0.091*** (0.022)	0.152*** (0.023)	-0.260*** (0.029)	0.141*** (0.022)
Year fixed effects	Yes	Yes	Yes	Yes
Evaluated peer controls	Yes	Yes	Yes	Yes
Dyad-level controls	Yes	Yes	Yes	Yes
Election-level controls	Yes	Yes	Yes	Yes
Candidate FE	No	No	Yes	Yes
Voter FE	Yes	Yes	No	No
R-squared	.2406169	.362537	.26115	.2944989
Log-likelihood	-95,485	-38,291.7	-111,934	-44,029.3
N	274,672	169,055	274,672	169,055

Note: Robust standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

T4.1 Different “control” groups

There are different types of members in the sample: members anticipating their own evaluation (“About-to-be candidate”), members that have been promoted (admins), members that are not going to be evaluated, and members that were candidates but have not been promoted. In order to rule out that our results are driven by comparing them to a group that could distort the results, we specifically compare members concerned with being evaluated to only two subgroups. For example, it could be possible that admins become less active after being elected. In this case the significant effect of “About-to-be candidate” is due to the activity reduction by promoted members and not by an increase in activity of members that want to be promoted.

Table T6 details the results of comparing our relevant group of members to elected admins (“Compared to promoted”) and members that have not been concerned (“Compared to before ATBC”). Whereas the comparison to elected admins supports our main analysis, the comparison to unevaluated users shows statistical significance only for H3. One possibility is the low number of observations (about 20% of the full sample) to explain the lack of significance, since the direction of the coefficient is the same and the size is similar. Another explanation could be that members that have are not about to be evaluated but might be aiming for adminship, appear to be similar in their strategy to avoid offending specific groups.

Table T6: Multinomial logistic regression, DV: Decision how to evaluate (abstention as base case), different comparison groups.

	Compared to admin			Compared to before ATBC		
	Neutral	Negative	Positive	Neutral	Negative	Positive
About-to-be candidate (ATBC)	1.262*** (0.120)	1.229*** (0.113)	1.133*** (0.104)	1.065*** (0.115)	1.015*** (0.115)	0.998*** (0.103)
Evaluation Pivotality	1.408*** (0.059)	1.830*** (0.058)	0.608*** (0.055)	1.492*** (0.100)	1.795*** (0.107)	0.669*** (0.080)
Candidate Activity	-1.093*** (0.303)	-1.120*** (0.250)	-1.026*** (0.194)	-0.949 (0.487)	-1.439*** (0.415)	-1.081*** (0.318)
ATBC * Evaluation Pivotality	-0.286* (0.120)	-0.660*** (0.110)	-0.237* (0.097)	-0.144 (0.136)	-0.316* (0.140)	-0.131 (0.112)
ATBC * Candidate Activity	0.208 (0.474)	-0.190 (0.397)	0.610 (0.346)	0.751 (0.535)	0.602 (0.491)	0.937* (0.402)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Evaluating member controls	Yes	Yes	Yes	Yes	Yes	Yes
Evaluated peer controls	Yes	Yes	Yes	Yes	Yes	Yes
Dyad-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Election-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared		0.364			0.383	
Log-likelihood		-158,793			-68,167.6	
N		182,032			80,479	

Note: Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

T5 Matching

T5.1 Robustness and description of PSM matching

Members can abstain and, thus, not participate in an evaluation. But we cannot observe whether members who did not participate were unaware of the RfA—for example, due to inactivity—or were aware and chose not to take part. We employ propensity score matching to ensure that all non-evaluations are likely made by members as an active decision to not evaluate a candidate and not because they were unaware of the election. Having constructed a credible counterfactual by calculating the propensity for each potential evaluation to result in an actual evaluation, we are able to select non-evaluations (“control”) that are similar to actual evaluations (“treatment”) in their observable factors. We select factors that we assume to be correlated with awareness of the RfA: namely the activity of the potential evaluator, the prominence of the candidate, and the similarity between evaluator and candidate (Table T7). We then used a logistic regression to determine how these factors explain whether potential evaluation would result in actual evaluations and used it to calculate the propensity for an evaluation.

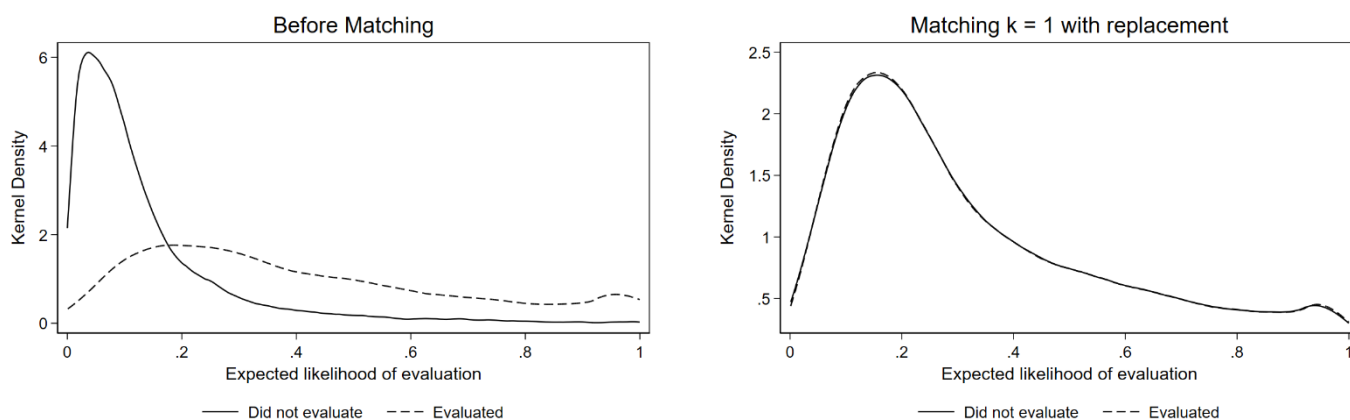
Table T7: Variables selected for propensity score matching

Evaluating member:	Tendency to evaluate (user fixed) Admin
Evaluating member and evaluated peer:	Edited pages Engagement Share of edits Tenure Number of elections participated
Dyad:	Work overlap

T5.2 Plots

The kernel density plots show the distribution for evaluations and non-evaluations. The second plot shows that the matching achieves a perfect overlay of an evaluation and non-evaluation. This is a very important feature of our paper, showing that the matching works well.

Figure T 5: Propensity scores before and after matching for (non)-evaluations.



T5.3 Comparison of samples

In addition to the density plots, we compare the means of the matched variables for both samples. Although some variables show a statistically significant difference, most differences are marginal and likely significant due to the large sample size. For example, if we randomly reduce the sample size to 20% of the original, most differences become insignificant. Furthermore, if we assume that the calculated propensity is a good predictor of “awareness,” the differences in individual variables do not matter.

Table T8: Comparison of samples

Variable	Non-eval. (Mean)	Evaluation (Mean)	Difference	p value
Evaluator-level controls				
Tendency to evaluate	0.126	0.119	0.007	0
Edited pages	7.187	7.222	-0.034	0
Engagement	0.034	0.078	-0.044	0
Share of edits	0.273	0.267	0.005	0
Tenure	23.137	22.236	0.901	0
No. of elections participated	89.206	86.993	2.213	0.029
Admin	0.34	0.382	-0.041	0
Candidate-level controls				
Edited pages	7.351	7.338	0.013	0.024
Engagement	0.005	0.036	-0.031	0
Share of edits	0.294	0.293	0.001	0.245
Tenure	23.938	24.161	-0.224	0.023
Previous attempts	0.374	0.381	-0.007	0.075
Dyad-level controls				
Work overlap	0.584	0.703	-0.119	0

T5.4 Alternative matching procedures

To test the robustness of our matching choices, we ran several alternative specifications. In Table T9, we run PSM with two neighbors ($k=2$), five neighbors ($k=5$), and one neighbor without replacement. In Table Y, we test the specification of one neighbor with replacement (same as in the main analysis) and vary the caliper (radius) between 0.01, 0.1, and 0.2. All results support our findings.

Table T9: Multinomial logistic regression, DV: Decision how to evaluate (abstention as base case), alternative matching

	k=2 (Two neighbors)			k=5 (Five neighbors)			No replacement		
	Neutral	Negative	Positive	Neutral	Negative	Positive	Neutral	Negative	Positive
About-to-be candidate (ATBC)	1.382*** (0.112)	1.337*** (0.100)	1.195*** (0.092)	1.407*** (0.110)	1.363*** (0.097)	1.217*** (0.090)	1.471*** (0.104)	1.422*** (0.093)	1.247*** (0.087)
Evaluation Pivotality	1.324*** (0.063)	1.688*** (0.050)	0.628*** (0.043)	1.326*** (0.061)	1.691*** (0.049)	0.627*** (0.042)	1.363*** (0.064)	1.735*** (0.047)	0.661*** (0.040)
Candidate Activity	-0.840*** (0.235)	-0.856*** (0.185)	-0.792*** (0.139)	-0.922*** (0.225)	-0.949*** (0.178)	-0.858*** (0.136)	-0.279 (0.214)	-0.326* (0.155)	-0.393*** (0.114)
ATBC * Evaluation Pivotality	-0.303** (0.111)	-0.614*** (0.099)	-0.403*** (0.085)	-0.365*** (0.108)	-0.677*** (0.096)	-0.460*** (0.082)	-0.450*** (0.102)	-0.769*** (0.090)	-0.526*** (0.077)
ATBC * Candidate Activity	-0.679 (0.437)	-1.190*** (0.355)	-0.182 (0.301)	-0.780 (0.461)	-1.269*** (0.375)	-0.271 (0.325)	-0.648 (0.380)	-1.143*** (0.314)	-0.063 (0.243)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluating member controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluated peer controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dyad-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Election-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared		0.346			0.346			0.350	
Log-likelihood		-239,301			-239,302			-235,719	
N		332,879			435,572			336,474	

Note: Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table T10: Multinomial logistic regression, DV: Decision how to evaluate (abstention as base case), alternative matching radii

	Radius = 0.01			Radius = 0.1			Radius= 0.2		
	Neutral	Negative	Positive	Neutral	Negative	Positive	Neutral	Negative	Positive
About-to-be candidate (ATBC)	1.487*** (0.105)	1.444*** (0.093)	1.274*** (0.087)	1.470*** (0.105)	1.426*** (0.093)	1.252*** (0.088)	1.464*** (0.104)	1.420*** (0.094)	1.254*** (0.088)
Evaluation Pivotality	1.364*** (0.052)	1.717*** (0.039)	0.676*** (0.035)	1.356*** (0.057)	1.714*** (0.043)	0.667*** (0.037)	1.361*** (0.059)	1.724*** (0.045)	0.665*** (0.038)
Candidate Activity	-0.484* (0.244)	-0.422* (0.170)	-0.501*** (0.120)	-0.435 (0.233)	-0.389* (0.164)	-0.467*** (0.117)	-0.293 (0.222)	-0.331* (0.161)	-0.412*** (0.115)
ATBC * Evaluation Pivotality	-0.435*** (0.103)	-0.764*** (0.094)	-0.532*** (0.079)	-0.383*** (0.101)	-0.768*** (0.091)	-0.525*** (0.078)	-0.435*** (0.102)	-0.777*** (0.092)	-0.529*** (0.078)
ATBC * Candidate Activity	-0.989* (0.431)	-1.537*** (0.342)	-0.296 (0.258)	-1.009* (0.421)	-1.373*** (0.335)	-0.170 (0.253)	-0.770* (0.391)	-1.313*** (0.329)	-0.197 (0.251)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluating member controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluated peer controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dyad-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Election-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared		0.338			0.340			0.344	
Log-likelihood		-196,774			-210,525			-221,192	
N		274,765			295,418			312,209	

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

T5.5 Alternative matching with coarsened exact matching (CEM)

To determine whether our results are driven by the specific matching algorithm (propensity score matching), we use a coarsened exact matching (CEM) as an alternative. A big difference is that CEM does not calculate a “propensity” but, depending on the values of selected criteria variables, it puts observations into groups with the same value range for the selected variables. We use the same variables that we used for the PSM method (Table T11). CEM can either automatically determine the “cut points” to put observations into different strata (‘brackets’) or users can manually define them. Again, the matching algorithm creates weights for observations if the strata they belong to do not have the same amount of treatment and control cases.

Table T11: Matching variables and cut points for CEM algorithm

Variable	Cut points Focal Member	Cut points Candidate
Evaluator-level controls:		
Tendency to evaluate	Automatic	
Admin	0/1	
Candidate-level and evaluator-level controls:		
Edited pages	Automatic	(0 4 6 8 10)
Engagement	(-1 -0.33 0.33 1)	2 automatic cut points
Share of edits	6 automatic cut points	5 automatic cut points
Tenure	Automatic	(0 12 24 36 48)
Number of elections participated	Automatic	(0 1)
Dyad-level controls:		
Work overlap	Automatic	

Comparing the means of both samples (evaluations and non-evaluations) (Table T12), most differences are small but statistically significant. The only notable differences are work overlap, the focal member’s tendency to evaluate, and the number of participated elections. We narrowed the cut points in an alternative specification and found non-significant differences. However, in that case, the number of active evaluations was much lower than in the propensity matching case and would not allow for a valid comparison across matching strategies.

Table T12: Comparison of samples created by coarsened exact matching

Variable	Non-eval. (Mean)	Non-eval. (SD)	Evaluation (Mean)	Evaluation (SD)	Difference	p-value
Evaluator-level controls						
Tendency to evaluate	0.124	0.147	0.047	0.091	0.077	0
Edited pages	7.181	1.355	6.992	1.417	0.189	0
Engagement	0.03	0.84	0.057	1.04	-0.027	0
Share of edits	0.272	0.114	0.262	0.117	0.01	0
Tenure	23.102	20.871	31.453	22.554	-8.351	0
No. elections participated	84.692	134.318	72.296	138.259	12.396	0
Admin	0.339	0.473	0.415	0.493	-0.076	0
Candidate-level controls						
Edited pages	7.354	1.281	7.355	1.276	-0.001	0.872
Engagement	0.006	1.002	0.024	1.023	-0.018	0.001
Share of edits	0.293	0.112	0.289	0.113	0.005	0
Tenure	23.903	20.381	23.547	20.116	0.356	0.001
No. previous attempts	0.368	0.801	0.379	0.804	-0.011	0.009
Dyad-level controls						
Work overlap	0.571	1.438	0.412	1.459	0.159	0

Overall, the regression results (Table T13) support our main findings, suggesting the same results as the PSM matching.

Table T13: Multinomial logistic regression, DV: Decision how to evaluate (abstention as base case), different CEM strategies compared.

	Test same weights			Test different weights		
	Neutral	Negative	Positive	Neutral	Negative	Positive
About-to-be candidate (ATBC)	1.625*** (0.116)	1.575*** (0.100)	1.357*** (0.084)	1.673*** (0.111)	1.591*** (0.095)	1.380*** (0.078)
Candidate Activity	1.462*** (0.101)	1.792*** (0.061)	0.616*** (0.060)	1.579*** (0.078)	1.864*** (0.051)	0.707*** (0.048)
Evaluation Pivotality	-0.551 (0.433)	-1.371*** (0.289)	-0.719** (0.236)	-0.168 (0.361)	-0.847*** (0.242)	-0.585** (0.181)
ATBC * Evaluation Pivotality	-0.430* (0.191)	-0.773*** (0.134)	-0.248 (0.128)	-0.595*** (0.158)	-0.869*** (0.120)	-0.379*** (0.103)
ATBC * Candidate Activity	-1.055 (0.609)	-1.131* (0.562)	-0.541 (0.475)	-1.075 (0.562)	-1.168* (0.476)	-0.382 (0.375)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Evaluating member controls	Yes	Yes	Yes	Yes	Yes	Yes
Evaluated peer controls	Yes	Yes	Yes	Yes	Yes	Yes
Dyad-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Election-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared		0.419			0.460	
Log-likelihood		-105,984			-95,436.6	
N		269,256			269,256	

Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

T6 Timing of evaluation decision

In the manuscript we highlight an additional analysis that investigates the timing of an evaluation. Here, we explain the logic and regressions of that analysis.

RfAs are supposed to be held over seven days, even though some might be extended or others are stopped prematurely (e.g., withdrawn nomination). In most cases the duration of an election is not a surprise to members and they can decide when to vote. Thus, we can categorize whether an evaluation was relatively early or late. We operationalize this by creating a dependent variable that captures the relative day of an evaluation, with 0 being the start of the election and 7 being the last day. We employ a Cox proportional hazards model to determine how much time passes before the event, an evaluation, occurs. Table T14 shows the full results of the predictors on the “hazard” of evaluating a candidate. Negative coefficients indicate a decreased risk in evaluation, i.e., an evaluation that, on average, will be later. For example, *About-to-be candidate* is positive, which means, on average, members that plan to be candidates will evaluate earlier. However, for negative evaluation (*ATBC * Negative Evaluation*) the coefficient becomes negative, indicating that members concerned with being evaluated delay their negative evaluation. Likewise, in RfA processes that are highly uncertain, members concerned about being evaluated wait before making an evaluation.

Table T14: Timing of evaluation, Cox proportional hazards Model, DV: Day of evaluation

About-to-be candidate (ATBC)	0.591 ^{***}	(0.047)
Negative Evaluation	1.438 ^{***}	(0.031)
Evaluation Pivotality	0.162 ^{***}	(0.021)
ATBC * Negative Evaluation	-0.672 ^{***}	(0.045)
Negative Evaluation * Evaluation Pivotality	-0.158 ^{***}	(0.033)
ATBC * Evaluation Pivotality	-0.156 ^{***}	(0.039)
Evaluating member:		
Activity	0.103	(0.131)
Edited Pages	-0.166 ^{***}	(0.037)
Edits	0.0998 ^{**}	(0.036)
Engagement	0.0142	(0.019)
Share of Edits	0.459 ^{***}	(0.138)
Tenure	-0.0103 ^{***}	(0.001)
No. of Elections Participated	-0.000599	(0.000)
Admin	-0.248 ^{***}	(0.042)
Relative Time to Vote	0.00926	(0.064)
Previous Failed Candidacy	0.429 ^{***}	(0.060)
Reversed Evaluations	0.00352	(0.002)
Deviation from Majority	-0.608 ^{***}	(0.070)
Evaluated peer:		
Edited Pages (Log.)	0.00713	(0.011)
Activity	-0.610 ^{**}	(0.068)
Edits (Log.)	-0.00154	(0.012)
Tendency to Evaluate Neg.	0.616 ^{**}	(0.213)
Engagement	-0.0114	(0.007)
Share of Edits	-0.219 ^{***}	(0.056)
Tenure	-0.00232 ^{***}	(0.000)
Edits Commented	-0.0731 [*]	(0.033)
Edits Reverted (Total)	0.000104 ^{**}	(0.000)
Edits Reverted (Share)	-0.115 ^{**}	(0.040)
Self-nomination	0.0383 ^{***}	(0.009)
Previous Failed Candidacy	0.0247 ^{**}	(0.009)
Activity Gaps	-0.0000878	(0.000)
Created Pages (Log.)	-0.00888	(0.007)
Average Similarity	-26.17 ^{***}	(1.651)
Dyad:		
Preceding Peer Approval	0.930 ^{***}	(0.034)
Preceding Peer Criticism	0.574 ^{***}	(0.091)
Work Overlap	0.0558 ^{***}	(0.005)
Edits of Candidate Deleted	0.265 ^{***}	(0.017)
Edits by Candidate Deleted	0.243 ^{***}	(0.015)
Edits of Cand. Deleted * Edits by Cand. Deleted	-0.0931 ^{***}	(0.011)
Election-level controls:		
Activity First Day of Election	0.00155 ^{***}	(0.000)
Current Approval Rating	0.743 ^{***}	(0.029)
Constant	-2.662 ^{***}	(0.122)
Year fixed effects	Yes	
N	954,706	

Note: Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

T7 Exclusion of alternative explanations

To determine whether role duality increases strategic behavior, we exploit a facet of the RfA process. We hand-coded all RfAs to determine whether a member self-nominated himself/herself or was nominated. Our rationale is that members who self-nominate are more likely driven by strategic behavior than members who were nominated by someone else while controlling for other factors as activity (Table T15). When we run sub-sample regressions with or without self-nominated members, we see that the interaction coefficients are stronger for self-nominated members (although the differences for Activity are not significant).

Table T15: Multinomial logistic regression; split sample: self-nomination

	No Self-nomination			Only Self-nomination		
	Neutral	Negative	Positive	Neutral	Negative	Positive
About-to-be candidate (ATBC)	1.336*** (0.118)	1.228*** (0.107)	1.127*** (0.097)	1.412*** (0.144)	1.352*** (0.129)	1.204*** (0.121)
Candidate Activity	-0.826** (0.293)	-1.091*** (0.229)	-0.909*** (0.169)	-0.921* (0.465)	-0.834* (0.347)	-0.847** (0.297)
Evaluation Pivotality	1.266*** (0.076)	1.747*** (0.058)	0.563*** (0.042)	1.682*** (0.071)	1.867*** (0.063)	1.047*** (0.057)
ATBC *Candidate Activity	-0.025 (0.543)	-0.669 (0.444)	0.252 (0.369)	-1.243 (0.781)	-1.257 (0.700)	-0.233 (0.594)
ATBC *Evaluation Pivotality	-0.043 (0.144)	-0.329** (0.122)	-0.237* (0.107)	-0.762*** (0.183)	-0.912*** (0.172)	-0.653*** (0.162)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Candidate-level control	Yes	Yes	Yes	Yes	Yes	Yes
Dyad-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Election-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared		0.353			0.324	
Log-likelihood		-156,948			-80,071.4	
N		193,852			80,845	

Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We also ruled out that a candidate's behavior can be explained by general conflict-avoidance or selection effects, i.e., only more engaged members become candidates. If that were the case, we would expect the same behavior to persist even after being elected. However, when using administrators as the main independent variable, the main effect and interaction become insignificant (Table T16).

Table T16: Multinomial logistic regression; 'Admin' as main independent variable

	Neutral	Negative	Positive
Admin	-0.042 (0.087)	-0.016 (0.073)	0.001 (0.062)
Candidate Activity	-0.858** (0.267)	-0.872*** (0.211)	-0.683*** (0.173)
Evaluation Pivotality	1.245*** (0.070)	1.476*** (0.060)	0.625*** (0.048)
Admin * Candidate Activity	-0.176 (0.338)	-0.264 (0.243)	-0.392* (0.197)
Admin * Evaluation Pivotality	0.156 (0.092)	0.391*** (0.084)	-0.132 (0.077)
Year fixed effects	Yes	Yes	Yes
Evaluator-level controls	Yes	Yes	Yes
Candidate-level control	Yes	Yes	Yes
Dyad-level controls	Yes	Yes	Yes
Election-level controls	Yes	Yes	Yes
Pseudo R-squared	0.345		
Log-likelihood	-239,523		
N	274,697		

Note: Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We also test whether our results can be explained that about-to-be candidates are better judges of quality. To do so, we created a measure of candidate quality. Based on all the new variables we have collected that describe the quality of a candidate, we ran a logistic regression predicting the likelihood of a candidate receiving positive evaluations based on candidate characteristics related to quality. This regression allowed us to calculate the "quality", a value between 0 and 1, of all candidates. To avoid endogeneity and simultaneity concerns, we only used the votes of elected members (admins). This way it is less likely that strategic behavior could bias the results. We use this propensity as a proxy for candidate quality. Alternatively, we measured the admin activity of a candidate post-RfA. Here, we reason that a low-quality candidate is likely not very active in his work as admin, and a high-quality candidate, on average, more active. We interact both indicators with our independent variable. If about-to-be candidates are better judges of quality, we should see them reject bad candidate and support good candidates more often. Table T17 shows that we only have partial evidence for this conjecture. About-to-be candidates tend to reject bad candidates more often (a significant reduction in the negative evaluation of higher quality candidates) but do not support high-quality candidates.

Table T17: Multinomial logistic regression; DV: decision how to evaluate, Moderator: Cand. Quality

	Admin Activity (Post-RfA)			Expected Candidate Quality		
	Neutral	Negative	Positive	Neutral	Negative	Positive
About-to-be candidate (ATBC)	1.688*** (0.166)	1.599*** (0.136)	1.172*** (0.131)	1.739*** (0.224)	1.989*** (0.223)	1.148*** (0.260)
Post Election Activity (Log.)	-0.003 (0.010)	-0.037*** (0.006)	0.042*** (0.005)			
(Expected) Candidate Quality				0.233 (0.344)	-0.709** (0.266)	-0.538 (0.290)
ATBC * Post Election Activity (Log.)	-0.077*** (0.020)	-0.093*** (0.017)	-0.019 (0.015)			
ATBC * (Expected) Candidate Quality				-0.639** (0.246)	-1.283*** (0.263)	-0.122 (0.307)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Candidate-level control	Yes	Yes	Yes	Yes	Yes	Yes
Dyad-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Election-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared	0.347			0.345		
Log-likelihood	-232,249			-239,502		
N	268,968			274,697		

Note: Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We then ran the H3 regression (Table 5 in the manuscript) with two subsamples separating high-quality and low-quality candidates (Table T18). If about-to-be candidates only focus on low-quality, inactive candidates, then we would expect the interaction coefficients of the low-quality candidates to be significantly higher. But that is not the case. To be fully sure, we ran both regressions in parallel as seemingly unrelated regressions and tested whether the interaction coefficients are significantly different. They are not.

Table T18: Multinomial logistic regression; Split sample: Candidate Quality

	High Cand. Quality			Low Cand. Quality		
	Neutral	Negative	Positive	Neutral	Negative	Positive
About-to-be candidate (ATBC)	1.281*** (0.112)	1.051*** (0.103)	1.106*** (0.095)	1.301*** (0.125)	1.193*** (0.106)	1.001*** (0.100)
Candidate Activity	0.963** (0.304)	0.305 (0.241)	0.234 (0.174)	-3.059*** (0.377)	-2.628*** (0.287)	-2.832*** (0.240)
ATBC * Candidate Activity	-0.338 (0.619)	-1.288* (0.524)	-0.060 (0.412)	-0.684 (0.626)	-1.053* (0.537)	0.071 (0.497)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Candidate-level control	Yes	Yes	Yes	Yes	Yes	Yes
Dyad-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Election-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared	0.375			0.299		
Log-likelihood	-128,097			-107,864		

N	169,150	105,547
---	---------	---------

Note: Robust standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

T8 Description of post-hoc analysis

The post-hoc analysis investigates the success of our hypothesized evaluation strategies. We select all candidates that have voted before their own RfA and have thus likely had strategic motives at some point and determine their approval rating (positive evaluations divided by positive and negative evaluations) and whether they have been approved as admins. Additionally, we determine the candidate-level controls analogous to the main analysis. Finally, we determine how often a candidate requested adminship (no. of repeated candidacies).

The summary statistics are shown in Table T19. The results for the main post-hoc analysis are in the manuscript (Table 6).

Table T19: Descriptive statistics for post-hoc analysis

		Mean	S.D.	Min	Max	0.	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	Elected	0.529	0.499	0.000	1.000	1.00											
(2)	Approval Rate	0.681	0.336	0.000	1.000	0.84	1.00										
(3)	Pos. eval. overall	8.342	16.750	0.000	207.000	0.06	0.15	1.00									
(4)	Neg. eval. overall	2.220	6.002	0.000	110.000	0.01	0.07	0.58	1.00								
(5)	Neg. eval. dummy	0.426	0.495	0.000	1.000	0.07	0.16	0.43	0.42	1.00							
(6)	Ratio of pos. eval.	0.811	0.240	0.000	1.000	0.07	0.06	0.05	-0.34	-0.56	1.00						
(7)	Difference from average tendency	0.353	0.163	0.000	0.500	0.04	0.00	-0.06	-0.44	-0.69	0.93	1.00					
(8)	Corr. neg. eval. and cand. centrality (H2a)	-0.019	0.123	-1.000	1.000	-0.04	-0.04	-0.06	-0.06	0.12	-0.02	0.00	1.00				
(9)	Corr. pos. eval. and cand. centrality (H2b)	0.069	0.152	-1.000	1.000	-0.02	-0.00	0.19	0.11	0.05	0.01	-0.01	-0.20	1.00			
(10)	Corr. neg. eval. and uncertainty (H3a)	0.040	0.133	-0.489	1.000	-0.03	-0.04	0.05	0.15	-0.01	-0.05	-0.08	0.26	-0.10	1.00		
(11)	Corr. pos. eval. and uncertainty (H3b)	0.004	0.142	-1.000	1.000	-0.11	-0.11	-0.01	-0.03	0.01	-0.02	-0.02	0.03	0.04	-0.05	1.00	
(12)	Average voting similarity with evaluators	0.381	0.486	0.000	1.000	-0.40	-0.49	-0.12	-0.07	-0.09	-0.06	-0.00	-0.01	0.01	0.00	0.08	1.00

T9 Test of multinomial categories

Using a multinomial logistic regression comes with several hypotheses that we need to test to test whether the model and the coefficient are fitting and relevant. First, we employ a Wald test of each coefficient across all four categories. A $p\text{-value} < 0.01$ indicates that the coefficients are significantly different across categories. Table T20 shows that this is the case.

Next, we test whether the addition or subtraction of any category would affect the result. The independence of irrelevant alternatives (IIA) assumption states that if one category is removed or included, this would not affect the coefficients of the other categories. Here, we use a seemingly unrelated regressions (*suest*) test. The *suest* test shows that all categories are significant, indicating that we cannot remove one category without altering the coefficients of the other categories.

Table T20: Independence of Irrelevant Alternatives Tests Using Seemingly Unrelated Regressions

	chi2	df	P>chi2
Abstention	4971.232	94	0
Neutral	1134.374	94	4.4e-178
Negative	4375.104	94	0
Positive	5489.469	94	0

Alternatively, we test whether any two categories can be combined without altering the overall results. This is likely if two categories are (almost) indistinguishable and combining two categories would be more efficient. The tests below (Table T21) show that this is not the case for any category pair.

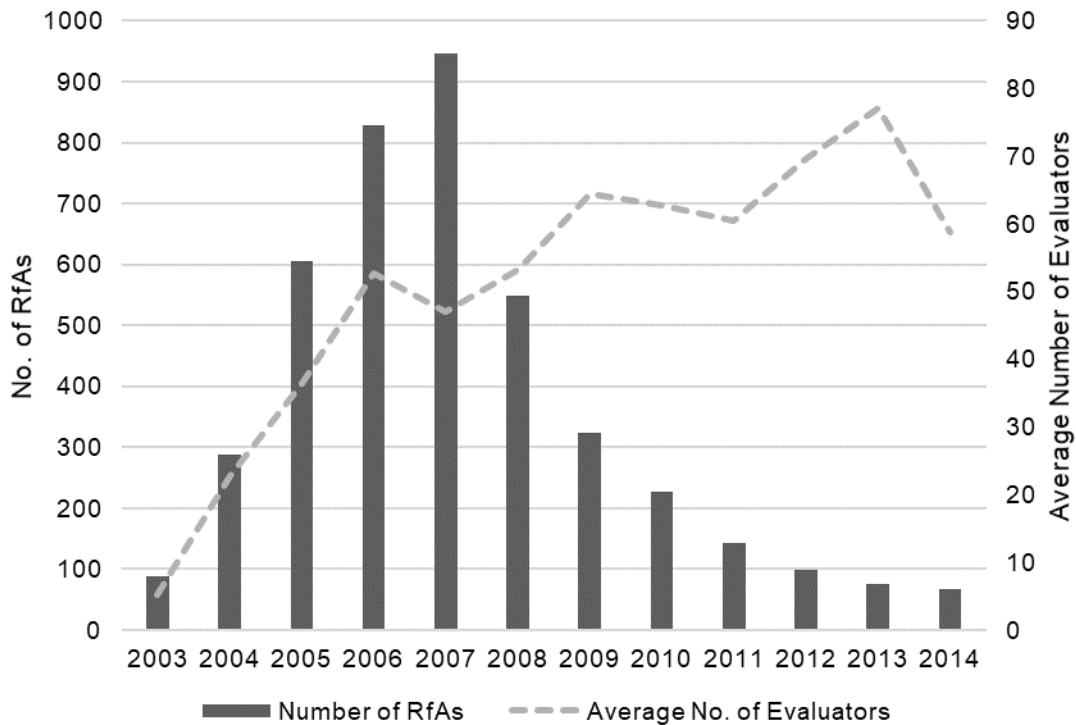
Table T21: Combination of Categories

	chi2	df	P>chi2
Abstention & Neutral	23576.18	46	0
Abstention & Negative	43329.49	46	0
Abstention & Positive	60722.59	46	0
Neutral & Negative	2904.667	46	0
Neutral & Positive	11395.48	46	0
Negative & Positive	36973.79	46	0

T10 Overview of the RfA process

This section provides a descriptive overview of the RfA process. Figure T6 shows the number of RfAs and average number of evaluators. There was a peak in number of RfAs in 2007. At that time, Wikipedia grew very quickly and attracted a lot of volunteers. The number of participants or evaluators per RfA has increased steadily since the beginning of RfAs.

Figure T6: RfAs and participation over time



The process of becoming an administrator on Wikipedia has remained fundamentally similar since the beginning of Wikipedia. Except for the first years, when admins were approved via email list, all users had to go through the same steps. Voters and candidates would go to the main RfA portal (https://en.wikipedia.org/wiki/Wikipedia:Requests_for_adminship) and find an overview of the process, current RfAs, and useful links.

To illustrate the process of an RfA, we have compiled screenshots from a successful Request for Adminship and annotated it. The original page can be accessed here:

https://en.wikipedia.org/wiki/Wikipedia:Requests_for_adminship/Xenocidic

Xenocidic [edit]

Final (72/13/2); Closed as successful by WJBscribe at 00:58, 4 June 2008 (UTC) [reply]

Xenocidic (talk · contribs) - This is one of the hardest RfA's that I've ever written, not because Xenocidic isn't deserving, but rather because I am being very hypocritical for writing it. You see, Xenocidic has not been actively editing for six months which is normally my criteria, but he's only a few weeks away from that magical six month mark. I thought about asking him to wait 2 weeks to run, that way I could say that he'd been active for six months, but I'm not sure what he might learn in two weeks that he doesn't already know. And it's not fair to him to force him to wait when he's ready now.

I even thought about serving as his admin coach---a program that I believe in---to get him past that six month mark. But to do so would be even more hypocritical than nominating him a few weeks early. Admin coaching, IMHO is *not a means to polish one's resume*, but rather a means to help somebody who needs some guidance. I've said elsewhere that I'd rather have a coachee that needs a lot of help, than the one who is all but ready to run. To me, coaching is a means to help guide a person to learn more about Wikipedia. It should not be something somebody does to get a check by their name---in fact, for some candidates coaching is, IMHO, a detriment. Xenocidic is one of those candidates.

1. The RfA process starts with the nomination by the user or by somebody else. The nomination is supposed to briefly summarize the reason why the candidate should pass the RfA.

Questions for the candidate [edit]

Dear candidate, thank you for offering to serve Wikipedia as an administrator. It is recommended that you answer these optional questions to provide guidance for participants:

1. What admin work do you intend to take part in?

A: At first, I will likely stick to what I know best, blocking persistent vandals reported at AIV, protecting pages that are being routinely vandalized, and deleting pages that clearly meet CSD criteria. I would also be adding myself to Category:Wikipedia administrators who will provide copies of deleted articles as it is my opinion that contributors should be granted the ability to take their work elsewhere or improve upon it such that it meets inclusion criteria. After completing Admin school, I would branch out and help with administrative backlog. I would, of course, continue to assist other users with any questions they may have as well as do my best to defuse conflicts between users.

2. What are your best contributions to Wikipedia, and why?

A: A lot of my work is somewhat behind the scenes and wiki-gnomish (copy-editing and such), but I think the most visible would be the major cleanups I completed on Realtime Worlds and WeMix.com (the latter still needs work but I saved it from a 'blatant advertising' CSD tag as the initial contributor had a conflict of interests, but I felt the article satisfied notability criteria). I also completed a major overhaul of Adopt-a-user, in particular helped to reduce the backlog of those seeking adoption from over 120 down to zero (this is starting to grow again, will have to do another adopter tree-shaking run). I do create articles whenever I'm the first on the scene with breaking news, as with Tom Clancy's H.A.W.X.. I'm also particularly pleased with two of the templates I designed - {{Truestatus}} & {{Statustop}} - which are used to effectively customize StatusBot and visually display said status, and allow other users to easily do so as well (for the latter I must credit Chetblong for the visual design). An overview of my contributions can be found at User:Xenocidic#Contributions. Lastly, I feel the interactions between myself and my adoptees has been helpful - for my adoptees and myself - as it has prompted me to expand my horizons.

2. After the nomination, the candidate responds to some standardized questions and questions asked by any member.

General comments [edit]

- See [Xenocidic's](#) edit summary usage with [mathbot's tool](#)[?]. For the edit count, see the [talk page](#).
- Links for Xenocidic: [Xenocidic](#) ([talk](#) · [contribs](#) · [count](#) · [AfD](#) · [logs](#) · [block log](#) · [lu](#) · [rfar](#) · [rfc](#) · [spi](#))

Please keep discussion constructive and civil. If you are unfamiliar with the nominee, please thoroughly review [Special:Contributions/Xenocidic](#) before commenting.

3. The RfA page provides quick access to the candidate's past activity. Users can easily can an overview what and how the candidate has interacted on Wikipedia.

Support [edit]

1. **Strong Support.** [Xenocidic](#) is an excellent editor with plenty of experience and has an excellent handle on policy. A fair number of his recent edits in the mainspace have been automated, but he has plenty of manual edits improving articles. Excellent communication skills and always civil; he's always willing to help other editors. My interactions with him have always been positive. Over 80 reports to AIV. I think he'll do a great job as an admin. [Useight](#) ([talk](#)) 00:58, 28 May 2008 (UTC) [reply]

Note: about 2 weeks ago he discovered huggle (or twinkle)... prior to that all of his edits were manual.[Balloonman](#) ([talk](#)) 01:08, 28 May 2008 (UTC) [reply]

Yeah, it's not a problem. I've used Huggle for a couple hundred edits myself and I don't mind an RFA candidate using automated tools, as long as they don't rely on it too much (see [User:Useight/RFA Standards](#)). Not a problem with Xenocidic. [Useight](#) ([talk](#)) 01:14, 28 May 2008 (UTC) [reply]

Yep - I've been testing out huggle the past week or so to see what all the fuss was about per the recent [discussion](#) at [WT:RFA](#). Extremely powerful and robust tool, kudos to Gurch (though I'm not sure yet if I'll continue using it). [xenocidic](#) ([talk](#)) 01:16, 28 May 2008 (UTC) [reply]

2. **per nom** [Dlohcierekim](#) 01:36, 28 May 2008 (UTC) [reply]
3. **Support.** I'm not so high on the social networking aspects of StatusBot, but I will not let that influence my decision making process here. I find the candidates' article building contributions sufficient, well rounded with contributions in other gnomish areas. He seems to be a good communicator, a solid base of policy knowledge and has a squeaky clean record deserving of community trust. [MrPrada](#) ([talk](#)) 02:17, 28 May 2008 (UTC) [reply]

4. Members can document their evaluation by writing on the RfA page and issuing their evaluation (support, oppose, or neutral) and optionally justifying their opinion.

Oppose [edit]

1. **Oppose**, in the answer to question 3, xenocidic refers to these 2 threads[1][2] (oldids[3][4]) where he tried to "act as a facilitator to resolve" a debate. From his comments there, it appears to me that xenocidic doesn't understand our **NPOV** policy, which aims to represent "fairly, and as far as possible without bias, all significant views that have been published by reliable sources." Here[5] he barges into the thread and declares "case closed", saying "can we put this one to rest." When someone disagrees he states "Please refer to the consensus/compromised already reached." [6]. A week later he's reverting[7][8] editors and telling them[9] on their talk pages that consensus had been reached (there wasn't any). If this editor thinks the way to resolve debates where one source says one thing and another source says another is to sweep in and say it's settled, I question their overall judgement when it comes to content disputes. Balloonman said xenocidic "only writes about video games" and that's also a concern of mine. --Pixelface (talk) 16:59, 28 May 2008 (UTC) [reply]

The consensus I was referring to is actually at the [Talk:GTA IV](#) (a majority of editors seemed to agree with the "unspecific Eastern European" compromise - Eastern Europe includes Serbia). However, I do agree that Pixel's suggestion [here](#) would be another way to resolve this particular dispute. [xenocidic \(talk\)](#) 17:18, 28 May 2008 (UTC) [reply]

2. **Oppose** — Individual fails to recognize that people can change their minds over time. [Kurt Weber \(Go Colts!\)](#) 19:21, 28 May 2008 (UTC) [reply]

umm... what? Would you elaborate somewhat please? [Thingg](#) 01:52, 29 May 2008 (UTC) [reply]

I think he means [10]. - [Bobet](#) 02:00, 29 May 2008 (UTC) [reply]

Yes, he's referring to the time I brought up his (admittedly ancient) self-nom after he opposed Guest9999's self-nom. [xenocidic \(talk\)](#) 02:04, 29 May 2008 (UTC) [reply]

Seems more like a personal beef ridden oppose than anything else. Sorry, I don't want to ABF. While I'm glad that, recently, Kurt has been elaborating more regarding his opposes, I wish it didn't have to smack of a potentially perceived slight or dig. That's the last I'll comment about this. [Wisdom89](#) 06:48, 29 May 2008 (UTC) [reply]

5. The negative evaluations (oppose) are documented below the positive evaluations. On average, the justification for (the relatively rare) negative evaluations are longer and lead to subsequent discussions (the indented comments).