Online Appendix for "Uninformed Voters and Corrupt Incumbents"

Contents

A1 Details of Corruption Cases

Table A1 gives the summary statistics of the corruption cases, whereas Tables A2 and A3 provide brief descriptions of each case used in the main analysis for the House and Senate, respectively.

Table A1: Summary statistics

	House 1968-2002	Senate 1968-2002	House Bank 1992
Total - identified	131	64	275
Total - retained	102	46	275
Retired	8 (7.8%)	8 (20.0%)	81 (29.5%)
Resigned	8 (7.8%)	1 (2.5%)	0 (0.0%)
Ran for reelection	73 (85.5%)	37 (80.4%)	194 (70.5%)
Reelected	59 (57.8%)	21 (45.7%)	176 (64.0%)
Merged onto ANES (obs)	30 (394)	17 (1,276)	77 (1,087)

The first row denotes all identified corruption cases. The second row denotes all cases following exclusion by the criteria mentioned in the paper. The percentages in the remaining rows show the shares in the amounts in the second row. The first and the third column are combined in the subsequent analysis. Sources: Brown (2006); Congressional Quarterly Almanac (Various); Congressional Quarterly (1992); Hirano and Snyder (2012); Noyer (1995); Roberds (1997)

Table A2: House scandals in the study (table split across pages)

Year	Representative	State	District	Description
1970	Martin McKneally	NY	27	Four-count indictment on charges of failing to file tax returns on income totaling $78,515
1975	Bill Nichols	AL	3	Pentagon investigation and reprimand by the Ethics Committee for gifts from Northrop while a member of the House Armed Service Committee.
1975	Robert Leggett	CA	4	Pentagon investigation and reprimand by the Ethics Committee for gifts from Northrop while a member of the House Armed Service Committee; Justice Dept. investigation for forging wife's signature when transferring property in Washington DC.

continued on next page

continued from previous page

Year	Representative	State	District	Description
1975	Dawson Mathis	GA	2	Pentagon and Ethics Committee investigation for gifts from Northrop while member of the House Armed Service Committee.
1975	John Flynt	GA	6	Ethics Committee investigation for avoiding property tax payment by transferring land to former aide, reclaiming land when taxes paid; free hunting trips from defense contractors.
1975	John Dingell	MI	16	Justice Dept. investigation for contributions and gifts from Gulf Oil lobbyists while drafting a law controlling oil spills.
1976	James Jones	OK	1	Guilty plea to a misdemeanor charge of not reporting a gift from a Gulf Oil lobbyist.
1976	Allan Howe	UT	2	Conviction on charges of soliciting for prostitution.
1978	Daniel Flood	PA	11	Indictment on perjury, bribery, and conspiracy charges; censure by the Ethics Committee.
1980	Frank Thompson	NJ	4	Conviction for soliciting a $50,000 bribe, 3-year sentence in ABSCAM.
1985	Tony Coelho	CA	15	Repaid funds raised above limit from S&L investor; investigation for insider trading.
1985	Michael Andrews	TX	25	Ethics Committee investigation of misuse of office for personal gain.
1987	Mario Biaggi	NY	19	8-year sentence, $242,000 fine for extortion, racketeering and conspiracy in the Wedtech case.
1987	Austin Murphy	PA	22	Reprimand by the Ethics Committee for hiring a "no-show" employee, misusing congressional property, and two counts of vote fraud.
1988	Charles Rose	NC	7	Reproach by the Ethics Committee for violations of campaign funds rules.
1988	Robert Garcia	NY	18	Conviction on charges of extorting $76,000 in payments disguised as consulting fees to his wife and a $20,000 interest-free loan from Wedtech.
1990	Floyd Flake	NY	6	Indictment on charges of diversion of $140,000 in funds to personal use; charges dropped in 1991 because of unfavorable testimony and restrictive rulings.
1992	Nicholas Mavroules	MA	6	Indictment on 17 counts of racketeering and extortion; guilty plea on 15 counts, 15-month sentence.
1992	House Bank Scandal			77 participants of the scandal.
1993	Melvin Reynolds	IL	2	Conviction for criminal sexual abuse, obstruction of justice and child pornography.

continued on next page

continued from previous page

Year	Representative	State	District	Description
1993	Dan Rostenkowski	IL	5	17-month sentence for misuse of office, misuse of franking privilege and diversion of funds.
1995	Newt Gingrich	GA	6	Citation by the Ethics Committee for failure to comply with House rules by allowing lobbyist to volunteer in his office.
1995	Gerald Kleczka	WI	4	3 arrests and a 30-day sentence on drunken driving charges.
1996	David McIntosh	IN	2	Admonition by the Ethics Committee for distributing documents created using an advocacy group's letterhead and misrepresentation of the Committee's findings.
1996	Richard Gephardt	MO	3	Rebuke by the Ethics Committee for failing to properly disclose income from a vacation property.
1996	Jim McDermott	WA	7	Justice Dept. investigation for making public an illegal tape recording of a conference call among House GOP leaders; conviction (see next).
1997	Jim McDermott	WA	7	Conviction of "willful and knowing misconduct" in illegal tapping, ordered to pay $1.05m to Rep. John Boehner in damages.
1998	Corrine Brown	FL	3	Ethics Committee investigation of influence peddling for release of a west-African businessman.
1999	Earl Hilliard	AL	7	Rebuke by the Ethics Committee for misuse of campaign funds for family business.
1999	Corrine Brown	FL	3	Ethics Committee investigation for influence peddling for release of a west-African businessman.
1999	Bob Barr	GA	7	FEC investigation of fund raising rules, including $100,000 in contributions exceeding the allowable limit.

Table A3: Senate scandals in the study (table split across pages)

Year	Senator	State	Description
1967	Thomas J. Dodd	CT	Censure by the Ethics Committee for misuse of political funds and double-billing for official and private travel.
1976	Vance Hartke	IN	Citation by the Ethics Committee for objecting to screening procedures that led to fines for 2 airlines, and excessive billing for foreign travel.
1976	J. Glenn Beall Jr.	MD	Violation of six sections of Maryland election laws by failing to report up to $200,000 in secret funds funneled by the White House.

continued on next page

continued from previous page

Year	Senator	State	Description
1976	Hubert Humphrey	MN	Part of the Justice Dept. investigation which resulted in two former campaign aides convicted for legal infractions.
1979	Edward Brooke	MA	Failure to disclose $49,000 in interest-free loans; Ethics Committee investigation announced, dropped following electoral defeat.
1979	Herman Talmadge	GA	Repaid Senate $37,125 in improperly claimed expenses; Justice Dept. investigation for misuse of campaign contributions and unreported taxes and gifts.
1980	Birch Bayh	IN	Ethics Committee finding of "neglect of duties" and Justice Dept. investigation for violation of franking privilege and perjurious statements in KoreaGate.
1986	Robert Kasten	WI	Arrest for drunken driving; Ethics Committee investigation for failure to file tax returns.
1986	Alfonse D'Amato	NY	Allegations of conflict of interest leading to the Ethics Committee investigation on 16 charges two years later.
1990	Phil Gramm	TX	Ethics Committee investigation for financial misconduct.
1991	Alfonse D'Amato	NY	Rebuke by the Ethics Committee on 16 charges of influence peddling and illegal contributions; connected to financial scandals of House members Biaggi and Garcia.
1992	John Glenn	OH	Rebuke by the Ethics Committee for "poor judgement" in the S&L scandal.
1992	John McCain	AZ	Rebuke by the Ethics Committee for "poor judgement" in the S&L scandal.
1992	Bob Packwood	OR	Ethics Committee eventually recommended expulsion for evidence tampering and perjury in a sexual harassment case.
1994	Kay Bailey Hutchison	TX	Indictment for misuse of state employees as state treasurer.
1994	Chuck Robb	VA	Justice Dept. investigation on charges of conspiracy and obstruction of justice in a phone tapping case against a longtime rival.
1998	Carole Moseley-Braun	IL	FEC investigation on accusation of misuse of $200,000 of leftover campaign funds for personal benefit.

Sources: Brown (2006); Congressional Quarterly Almanac (Various); Congressional Quarterly (1992); Hirano and Snyder (2012); Noyer (1995); Roberds (1997).

A2 Construction of the Political Awareness Score

As described in the text, the political awareness measure is constructed based on a number of items from the biennial American National Election Study (ANES) time-series surveys conducted between 1968 and 2002. The items probe respondents' factual knowledge of politics and institutions, such

as the party holding the majority in the House, their ability to recognize political figures (such as the name of the Vice President), and the proper placement of parties and candidates on a left-right scale on policy issues such as health care and defense. The questions are coded so that correct answers receive a positive value, whereas incorrect answers receive a value of zero. All missing values indicating the inability to respond are coded as incorrect (categories such as "DK", "Cannot judge", and "Did not rate").[1] The remaining missing values are multiply imputed along with missing values in the other respondent-level variables used in the analysis (see Section A4 below for more details). The full list of items used in the construction of the measure for each survey is available upon request.

Unlike other studies, which typically construct an additive scale with a varying maximum value, I use exploratory factor analysis on a single factor to build an integral scale. I build the scale for each survey separately, as the type and difficulty of the questions are not constant over time.[2] The value of the scale obtained from factor analysis for each respondent is then reexpressed in terms of the rank of that respondent's score with respect to the score of all other respondents in the survey. Ties are given an average rank (e.g. two observations with the same awareness score that would get a unique rank of 2 and 3 receive a rank of 2.5). Then, the ranks are divided by the highest score within each survey, and so the measure represents the ranked score with respect to the maximum awareness. Such procedure makes the measure comparable across surveys.[3] The average reliability score, as measured by the Chronbach's α, for the measure is .86.[4]

A3 1992 ANES House Bank Survey Items

In the 1992 American National Election Study, respondents were asked whether they had heard of the 1992 House Bank scandal (item v925715), knew if their representative wrote bad checks (v925718), and if so, whether they had written a few or a lot of them (v925719), and whether overdrafts were unlawful (v925720). The survey also elicited respondents' attitudes towards the egregiousness of the representatives' check writing (v925716). To build the appropriate variables used as outcomes in Figure 3 in the text, I cross-checked the respondents' answers with the data on the actual check-writing by individual legislators.[5] Correct answers receive the value of 1, and

[1] Mondak 1999; 2001 and Mondak and Davis (2001) condemn such practice, arguing that "don't knows" conceal partial knowledge. The authors recommend assigning any such persistent answers to substantive response categories at random, mimicking random guessing. However, such a procedure has been shown unwarranted on ANES and experimental data in a series of papers thereafter (Bennett, 2001; Luskin and Bullock, 2004 2011; Sturgis, 2006; Sturgis, Allum, and Smith, 2008).

[2] A typical scale includes around fifteen items, and more than one factor is usually found. However, the factor loading on the first factor is always noticeably stronger than the other factors. Item weights are very similar for similar or identical items across surveys, giving confidence that questions have a stable ability to measure latent political awareness.

[3] Normalizing to mean zero and standard deviation of one is another option, but this measure is somewhat misleading because factual items are not of constant difficulty from survey to survey: normalized measure ranges from less than 2 to more than 3 standard deviations on each side of the mean in different surveys. Nonetheless, the correlation between the normalized score and the rank score is greater than .99, and the results with the normalized score are very similar.

[4] As noted by Zaller (1992, p.337), scales with alpha reliability below .8 can fail to detect non-monotonicity or effects of attitude change, both of which may have important consequences for the analysis.

[5] Overdraft data are taken from "Voters Enraged Over House Bank Abuses." CQ Press Electronic Library, CQ Almanac Online Edition, `http://library.cqpress.com/cqalmanac/cqal92-1106904`.

incorrect the value of 0.[6]

The estimates shown in Figure 3 in the text derive from separate OLS regressions of each binary variable on political awareness, an indicator of whether a respondent's representative was involved in the scandal, and a number of control variables: incumbent's party affiliation (Democrat = 1), respondent's age, gender, race, indicator for urban area, region, family income indicators, respondent party ID, strength of partisanship, retrospective economic evaluation, presidential approval, ideological distance from the incumbent, logged vote margin for the incumbent from the previous election cycle, incumbent seniority, and campaign media intensity (see Section A5 below for details on the variable coding).[7]

A4 Multiple Imputation

Since ANES suffers from considerable item-non response, I multiply impute the data. Discarding missing data induces inefficiency and possibly bias (Rubin, 1987). Multiple imputation alleviates the loss of efficiency. It also eliminates bias if the responses are missing at random (MAR), conditional on all the variables included in the imputation stage. MAR is not empirically verifiable, but it is more plausibly satisfied when many relevant predictors of the response are included (Rubin, Stern, and Vehovar, 1995; Gelman and Hill, 2007). Therefore, in the imputation stage I include a number of variables in addition to the ones used in the final analysis. Whenever available, I include survey design variables. I further include variables that have been shown to correlate well with political sophistication (Zaller, 1992): education, media use, political participation, and interest in campaigns. I also include a number of demographic characteristics that have been shown to interact with information levels (Bartels, 1996) and to summarize well the information in ANES (Sekhon, n.d.). Finally, I include the counts of likes and dislikes of the major parties and congressional candidates for better imputation of the measures of incumbent support. Imputations are weighted with the ANES-provided sampling weights, which are also used in the analysis stage.

I use the chained-equation multiple imputation method (van Buuren, Boshuizen, and Knook, 1999; Royston, 2004). It is flexible, particularly for imputation of categorical variables and interval-censored variables such as the feeling thermometer scores (Royston, 2007). Imputations are performed using the package `ice` in `Stata 12` (Royston, 2004 2005). In many instances, three to five imputed datasets are sufficient for large efficiency gains (Rubin, 1987). Item non-response, however, is high in ANES, particularly for incumbent support items, and so I opt for fifteen imputations, balancing the validity of the imputed values and the size of the data to be imputed. Separate imputations are performed for the House and for the Senate, even though the set of variables is almost identical. The reason is that in many cases values of the variables are incompatible between the two datasets.

All the missing values for all the variables for all respondents in the ANES are imputed. This

[6] For the question on whether check-writing was unlawful, the correct answer for the majority of the cases is that it was *not* unlawful. However, 22 current and former members were cited by the House Ethics Committee as having abused their privileges at the House Bank, and several members subsequently faced felony charges and convictions. For these cases, the correct answer would be that the act was unlawful, and I code it accordingly. For the item probing whether a respondent's representative wrote a lot of checks or a few, I take ten checks as the cut-off point. Obviously, this rule is arbitrary. The findings for this measure are only suggestive. I tried different values of more than 25, 50, 75, and 100 checks as well. The results are similar, and I choose to report the results of this specification because it balances the number of cases in both ("a lot" as well as "a few") categories.

[7] The full set of estimates is available upon request.

means that the missing candidate preference of nonvoters is imputed along with the missing candidate preferences of voters. It also means that the missing values for turnout are imputed. In the main analysis, I restrict the sample only to those respondents for whom the value of the turnout variable equals one – imputed or not. However, the results are substantively quite similar if I run the analysis on the entire sample (i.e. without restricting the sample to voters).

A5 Coding of Covariates and Summary Statistics

Summary statistics for the key variables for the House and Senate are given in Tables A4 and A5, respectively. The number of observations reflects the 15 imputed datasets.

Table A4: Summary statistics for key variables, House

	Obs.	Mean	St. dev	Min	Max
Pol. awareness (rank)	437610	0.50	0.29	0	1
Scandal	437610	0.05	0.22	0	1
Incumbent vote	437610	0.71	0.45	0	1
Incumbent FT score	322455	62.31	23.00	0	100
Incumbent approval	273465	0.82	0.38	0	1
Incumbent PID (Dem.)	437610	0.60	0.49	0	1
Redistricted	437610	0.05	0.21	0	1
PID	437610	0.39	2.02	-3	3
Econ. eval	437610	0.07	0.80	-1	1
Coattails	437610	0.50	0.50	0	1
Distance	437610	1.89	1.33	0	6
Vote margin (log)	437610	3.36	1.03	-2	5
Tenure (log-yrs)	437610	1.95	0.81	0	4
Campaign intensity (resid.)	437610	0.01	0.24	-1	1

Respondents' party identification relative to the incumbent and the self-reported strength of partisanship are combined into one variable, so that strong co-partisans (respondents with the same party ID as the incumbent's party affiliation) receive the value of 3, independents the value of 0, strong out-partisans the value of -3, and weak and leaning co- and out-partisans the remaining values in-between.

In Tables A6 and A7, I add a number of additional variables to the main specification. Ideological distance from the incumbent is an absolute distance between the self-reported placement of the respondent on the left-right ideological scale and the DW-nominate first dimension of each incumbent from Poole and Rosenthal (2007). Both measures are standardized at mean zero and standard deviation one, reexpressed on a scale $-3/3$ in discrete steps of size one, and their absolute difference is taken. Economic evaluation is coded from the retrospective "pocketbook" item with the value of 1 if a respondent's economic situation is perceived as "better," 0 if "the same," and -1 if "worse than a year ago." Binary variable for the presidential approval is constructed so that (dis)approval of the President of the same party as the incumbent receives the value of 1, and 0 otherwise, thus representing approval "coattails."

Table A5: Summary statistics for key variables, Senate

	Obs.	Mean	St. dev	Min	Max
Pol. awareness (rank)	250680	0.50	0.29	0	1
Scandal	250680	0.08	0.27	0	1
Incumbent vote	250680	0.60	0.49	0	1
Incumbent FT score	155820	59.00	24.33	0	100
Incumbent PID (Dem.)	250680	0.53	0.50	0	1
PID	250680	0.11	2.04	-3	3
Econ. eval	250680	0.06	0.80	-1	1
Coattails	250680	0.48	0.50	0	1
Distance	250680	1.94	1.34	0	6
Vote margin (log)	250680	2.44	1.19	-1	5
Tenure (log-yrs)	250680	2.21	0.67	-2	4
Campaign intensity (resid.)	250680	0.01	0.19	-1	1

Characteristics of the incumbent are incumbent party affiliation, seniority, and the vote share in the previous election cycle. Incumbent party ID is a binary indicator taking the value of 1 if the incumbent is a Democrat, and 0 otherwise. I transform the two-party vote share from the previous cycles into a logged vote margin. Seniority is expressed in log-years. I also include a measure of the intensity of the electoral campaign. In the absence of a more suitable measure, I follow Zaller (1992) by utilizing the ANES item(s) tapping into the exposure to the campaign through the print media. This variable is averaged at the district or state level, and then purged of its correlation with political awareness and any temporal effects, by regressing the average exposure to the media on political awareness and a set of year dummies. Political awareness significantly and positively predicts self-reported exposure to campaigns. To verify the validity of this purged measure, I test the null hypothesis that campaign intensity is the same in races with a corrupt incumbent compared to races with a clean incumbent. The null hypothesis is rejected at $p < .001$ on a two-tailed mean-comparison test, and as expected, residuals are larger in scandal races.

A6 Coefficient Estimates and Robustness Checks

Table A6 gives the parameter estimates for the House from the main model (column 1), and from additional specifications which include a number of additional control variables (column 2-4). Table A7 gives the same results for the Senate regressions. In each column, the top panel shows the difference in the predicted probability of incumbent vote for corrupt relative to clean incumbents, at five values of political awareness: 1st, 25th, 50th, 75th, and 99th percentile. These differences represent a subset of values plotted in Figure 1 in the text. The bottom panel shows the coefficient estimates and standard errors. To save on space, the bottom panel excludes the estimates for the election fixed effects.

Table A8 shows that the results are not sensitive to an alternative definition of corruption. The first and third columns show the estimates from the main model used in the analysis in the paper, for the House and Senate, respectively. The second and fourth columns broaden the

Table A6: Additional control variables – House

	(1)		(2)		(3)		(4)	
	Support for corrupt vs. clean incumbent							
1st pctile	0.025	0.045	0.025	0.047	0.028	0.045	0.049	0.049
25th pctile	-0.006	0.034	-0.011	0.034	-0.008	0.032	0.010	0.037
50th pctile	-0.038	0.028	-0.049*	0.026	-0.047*	0.024	-0.032	0.030
75th pctile	-0.072**	0.031	-0.086***	0.028	-0.084***	0.027	-0.071**	0.032
99th pctile	-0.104***	0.040	-0.122***	0.038	-0.121***	0.038	-0.109***	0.042
	Coefficient estimates							
Pol. Awareness	-0.015	0.014	0.024	0.059	0.066	0.060	0.065	0.060
Pol. Awareness × scandal	-0.133*	0.068	-0.152**	0.069	-0.152**	0.069	-0.152**	0.070
Scandal	0.029	0.047	0.029	0.047	0.029	0.047	0.054	0.092
PID	0.102***	0.002	0.081***	0.005	0.095***	0.009	0.096***	0.009
Dem. inc.			0.007	0.022	0.011	0.022	0.010	0.022
Dem. inc. × aware.			0.004	0.032	-0.005	0.032	-0.006	0.032
Redist.			-0.031	0.021	-0.032	0.021	-0.034	0.021
Econ. eval.			0.016	0.011	0.014	0.013	0.014	0.013
Coattails			0.073***	0.021	0.087***	0.022	0.086***	0.022
Distance			0.002	0.006	0.000	0.006	-0.000	0.006
Vote margin			0.030***	0.010	0.031***	0.010	0.032***	0.010
Tenure			-0.006	0.013	-0.006	0.013	-0.007	0.013
Intensity			-0.006	0.036	0.001	0.036	0.007	0.037
Vote margin × aware.			0.006	0.015	0.004	0.015	0.004	0.015
Tenure × aware.			-0.003	0.020	-0.005	0.020	-0.004	0.020
Econ. eval. × aware.			-0.008	0.017	-0.004	0.019	-0.004	0.019
Coattails × aware.			0.037	0.031	0.036	0.034	0.036	0.034
PID × aware.			0.006	0.008	-0.019	0.014	-0.020	0.014
Distance × aware.			-0.042***	0.010	-0.040***	0.010	-0.040***	0.010
Intensity × aware.			0.011	0.055	0.008	0.055	0.008	0.055
Econ. eval × PID					-0.001	0.006	-0.001	0.006
Coattails × PID					-0.032***	0.009	-0.033***	0.009
Distance × PID					0.000	0.003	0.000	0.003
Econ. eval × aware. × PID					-0.002	0.009	-0.002	0.009
Coattails × aware. × PID					0.009	0.014	0.009	0.014
Disatance × aware. × PID					0.011**	0.005	0.011**	0.005
PID × scandal							-0.005	0.010
Econ. eval × scandal							-0.008	0.021
Coattails × scandal							0.036	0.035
Distance × scandal							0.006	0.012
Vote margin × scandal							-0.011	0.023
Tenure × scandal							-0.007	0.025
Intensity × scandal							-0.107	0.085

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

definition of corruption to include not just the cases with investigative action, but any corruption scandal identified in the sources I rely on (Brown, 2006; Congressional Quarterly Almanac, Various; Congressional Quarterly, 1992; Hirano and Snyder, 2012; Noyer, 1995; Roberds, 1997). Once again, the top panel shows the difference in the predicted probability of support for corrupt vs. clean incumbents for five levels of political awareness (this is equivalent to the quantities plotted shown in Figure 1 in the paper). The bottom panel shows the point estimates. The results are quite

Table A7: Additional control variables – Senate

	(1)		(2)		(3)		(4)	
	Support for corrupt vs. clean incumbent							
1st pctile	0.035	0.049	0.042	0.055	0.040	0.052	0.018	0.057
25th pctile	0.006	0.038	0.012	0.043	0.011	0.040	-0.010	0.044
50th pctile	-0.024	0.031	-0.022	0.034	-0.020	0.032	-0.037	0.034
75th pctile	-0.056*	0.033	-0.052	0.033	-0.050	0.031	-0.064**	0.032
99th pctile	-0.088**	0.041	-0.083**	0.039	-0.080**	0.039	-0.091**	0.037
	Coefficient estimates							
Pol. Awareness	0.011	0.022	0.033	0.080	0.051	0.081	0.044	0.080
Pol. Awareness × scandal	-0.126*	0.069	-0.128*	0.067	-0.123*	0.067	-0.111	0.069
Scandal	0.039	0.051	0.044	0.054	0.042	0.054	0.319***	0.106
PID	0.122***	0.003	0.087***	0.007	0.095***	0.011	0.095***	0.011
Dem. inc.			0.071**	0.030	0.070**	0.030	0.071**	0.030
Dem. inc. × aware.			-0.046	0.038	-0.046	0.039	-0.044	0.039
Redist.			0.006	0.017	0.006	0.017	0.010	0.017
Econ. eval.			0.105***	0.028	0.107***	0.028	0.110***	0.028
Coattails			0.001	0.009	0.001	0.010	0.002	0.010
Distance			0.010	0.012	0.010	0.012	0.011	0.012
Vote margin			0.016	0.019	0.016	0.019	0.022	0.019
Tenure			0.040	0.074	0.040	0.075	0.043	0.075
Intensity			0.029*	0.016	0.027*	0.016	0.027*	0.016
Vote margin × aware.			-0.014	0.026	-0.015	0.025	-0.011	0.025
Tenure × aware.			-0.005	0.024	-0.005	0.024	-0.006	0.024
Econ. eval. × aware.			0.042	0.042	0.047	0.042	0.044	0.042
Coattails × aware.			0.027***	0.009	0.027	0.016	0.026	0.016
PID × aware.			-0.026*	0.015	-0.026*	0.015	-0.027*	0.015
Distance × aware.			-0.143	0.106	-0.141	0.108	-0.138	0.108
Intensity × aware.					-0.002	0.006	-0.002	0.006
Econ. eval × PID					-0.018	0.012	-0.018	0.012
Coattails × PID					0.001	0.004	0.001	0.004
Distance × PID					-0.004	0.009	-0.004	0.009
Econ. eval × aware. × PID					-0.007	0.019	-0.006	0.019
Coattails × aware. × PID					0.001	0.005	0.001	0.005
Disatance × aware. × PID							0.005	0.011
PID × scandal							-0.040*	0.021
Econ. eval × scandal							-0.018	0.039
Coattails × scandal							0.001	0.015
Distance × scandal							-0.034	0.024
Vote margin × scandal							-0.091***	0.030
Tenure × scandal							-0.111	0.205
Intensity × scandal								

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

similar with this broader definition of corruption.

As mentioned in the paper, the dependent variable used in the main analysis is potentially problematic because voting for the election winner is overreported (Wright, 1993), and incumbents are overwhelmingly the winners. Here, I show that the results are quite similar when alternative measures of incumbent support that may be less sensitive to over-reporting are used.

As alternative measures, I use incumbent approval (only available for the House), and incumbent

Table A8: Results with alternative corruption samples

	House		Senate	
	Main Sample	All Scandals	Main Sample	All Scandals
	Support for corrupt vs. clean incumbent			
1st pctile	0.025	-0.002	0.037	-0.007
	(0.045)	(0.043)	(0.051)	(0.045)
25th pctile	-0.006	-0.027	0.008	-0.028
	(0.034)	(0.031)	(0.039)	(0.035)
50th pctile	-0.038	-0.055**	-0.024	-0.048*
	(0.028)	(0.025)	(0.031)	(0.028)
75th pctile	-0.072**	-0.082***	-0.055*	-0.069**
	(0.031)	(0.027)	(0.032)	(0.027)
99th pctile	-0.104***	-0.107***	-0.086**	-0.089***
	(0.040)	(0.036)	(0.041)	(0.032)
	Coefficient estimates			
Pol. Awareness	-0.015	-0.016	0.011	0.012
	(0.014)	(0.014)	(0.022)	(0.023)
Pol. Awareness × scandal	-0.133*	-0.109*	-0.126*	-0.082
	(0.068)	(0.061)	(0.069)	(0.058)
Scandal	0.029	0.000	0.039	-0.008
	(0.047)	(0.042)	(0.051)	(0.047)
PID	0.102***	0.103***	0.122***	0.121***
	(0.002)	(0.002)	(0.003)	(0.003)
Constant	0.616***	0.616***	0.545***	0.546***
	(0.023)	(0.023)	(0.046)	(0.046)
N	18,996	18,996	11,060	11,060
Scandal N	1,481	1,711	1,276	1,818

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

feeling thermometer scores. The former is a standard binary measure; the latter lies in the $[0, 100]$ interval.[8] Figure A1 shows the same results as in Figure 1 in the paper, but with stated vote choice replaced with these alternative measures of incumbent support. The results are quite similar to those for incumbent vote.

A7 Strategic Engagement in Corruption and Attenuation Bias

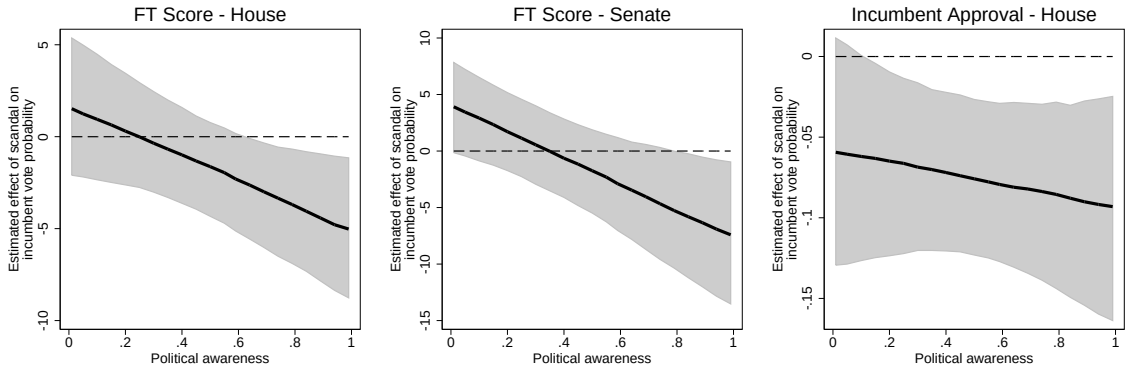
In the text, I argue that an important non-random aspect of the data-generating process – politicians' potential strategic engagement in corruption in response to their electorates' political awareness – likely induces attenuation bias in rejecting the null hypothesis. Here, I provide details on the reasoning behind this statement.

Figure A2 shows the extensive-form game tree of the formal model mentioned in the text. The model is a simple complete-information game with two actors: a representative voter and an incumbent politician. The voter can have one of two levels of political awareness: low or high.[9]

[8] It would be better to use the difference between incumbent and challenger FT scores, rather than relying on incumbent FT score alone, since the difference helps account for interpersonal differences in scoring. However, the missingness rate for the challenger FT scores is very high, making the difference unreliable.

[9] This binary awareness space is used for simplicity; the results would be substantively unchanged if the awareness

Figure A1: Alternative measures of incumbent support



The lines represent the difference in the predicted probability of voting for a corrupt incumbent relative to a clean incumbent for different levels of political awareness. Results are based on the model equivalent to the model in equation 1, except for the dependent variables, shown in each panel of the figure. Negative values imply that the predicted vote probability for a corrupt incumbent is lower than for a clean incumbent at a given level of political awareness. Estimates are calculated for values of political awareness from the 1st percentile to the 100th percentile in five-percentile steps. The shaded area represents the 90 percent confidence interval. All quantities are averaged over fifteen imputed datasets.

The voter makes a binary decision: to reelect the incumbent (E) or not ($\neg E$). The incumbent can be either involved in a corruption scandal, or not involved. The voter observes with certainty if the incumbent had engaged in corruption. The incumbent also makes a binary decision: whether to run for reelection (R) or not ($\neg R$).

The timing of the game is as follows. In this model, the decision of a politician to engage in corruption is not related to political awareness of the electorate. Therefore, nature assigns corruption, with probability r.[10] Next, nature assigns the incumbent to a high-awareness electorate (represented by a representative voter) with probability p.[11] The incumbent subsequently decides whether to run for reelection or not. If the incumbent does not run, the game ends. If the incumbent runs, the voter then decides whether to reelect the incumbent or not, and the game ends.

The preferences of the actors are as follows. I assume that H_{A1} is true, i.e. that a high-awareness voter is less likely to vote for a corrupt incumbent than a low-awareness voter. In line with H_{A1}, the high-information voter prefers to reelect the incumbent only if the incumbent is not engaged in corruption; the low-information voter always reelects the incumbent, irrespective of whether the

space was less coarse.

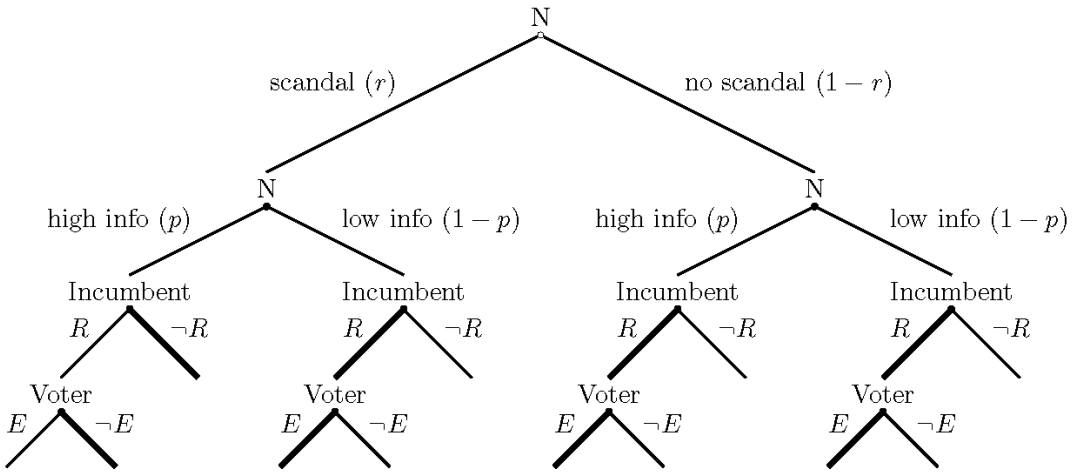
[10]Note that this assignment need not be interpreted as random from the perspective of the voter; it is simply random conditional on political awareness, and the remaining potential factors influencing the decision to engage in corruption are left unmodeled.

[11]Assignment probabilities are stated simply as part of the convention, but are not of interest in this modeling exercise.

incumbent is corrupt or not.[12] Running for reelection is costly, and so the incumbent prefers not to run than to run and lose. However, winning reelection outweighs the cost of running, and so the incumbent always prefers to get reelected.

Based on this simple structure, the only strategic actor in the game is the incumbent, who chooses whether to run for reelection based on the assignment of corruption and the voter preferences. Since the game is complete-information, the equilibrium concept is the subgame perfect Nash equilibrium (SPNE), obtained by the simple application of backward induction. I do not impose a specific payoff structure, as any utility function which satisfies the preferences of the actors solves for the same equilibrium of the game.

Figure A2: First model – reelection bid is strategic



In Figure A2, thick lines represent equilibrium play based on backward induction. Only the corrupt incumbent facing the high-awareness electorate chooses not to run for reelection. The implications for the empirical identification of the results are discussed in the paper. While the resulting equilibrium is start (i.e. no corrupt incumbent facing the high-awareness electorate runs for reelection), this is due to the simplifying assumptions made. But the general result, that strategic politicians' behavior likely induces attenuation bias would remain in a model with similar assumptions that would yield more realistic predictions.[13]

Figure A3 shows the extensive-form game tree of a similar model, but one where it is assumed that in addition to retirement, an incumbent's decision to engage in corruption is also strategic. The actors are the same as in the first model. The voter and the politician also have the same characteristics: awareness levels and involvement in scandal, respectively. The voter makes the same binary decision whether to reelect the incumbent. The politician's choice set, however, is

[12]The low-awareness voter need not always prefer to reelect the incumbent. It is sufficient to assume that she has a higher preference for the incumbent than the high-awareness voter.

[13]Political awareness of the electorate likely affects what kind of politicians select to run for office, not just whether an incumbent retires. I abstract from such type of political selection here, but it is easy to imagine that it only induces stronger attenuation bias. Namely, "bad" politicians, i.e those inherently more prone to corruption, would be less likely to run for office in high-awareness districts, precisely where according to H_{A1} voters are more likely to punish corruption.

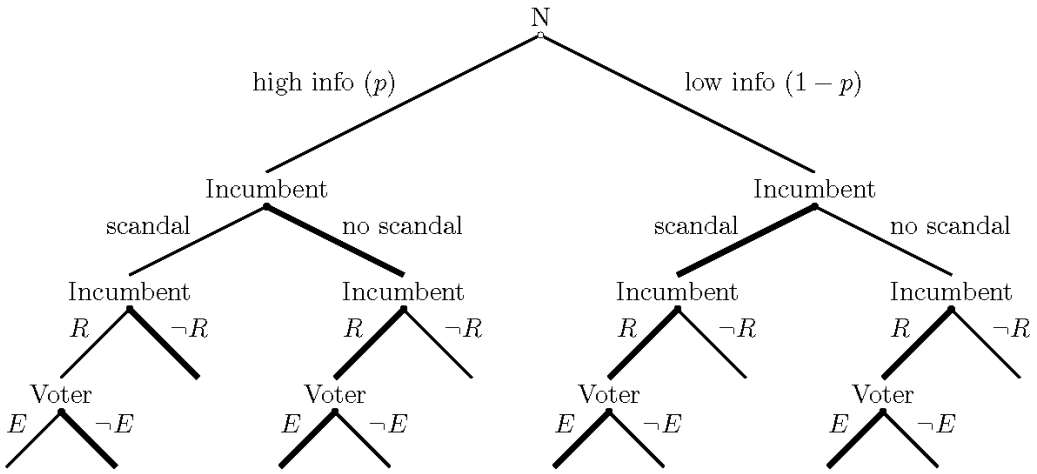
changed, because in this model I assume that both the decision to engage in corruption and to run for reelection depend on the extent to which voters are informed about politics.

The timing of the game is therefore slightly different from the first model. Nature moves first, but only to assign the incumbent to a high- or low-awareness electorate. Next, the incumbent chooses whether to engage in corruption ("scandal") or not, and then whether to run for reelection.[14] The subsequent play is the same as in the first model.

The preferences of the voter are the same as in the first model. The incumbent prefers engaging in corruption and getting reelected to being clean and getting reelected, and being clean and getting reelected to being engaged in corruption and voted out of office. This preference ordering is consistent with corruption benefits being higher than costs of engaging in corruption (such as reputation concerns), but the net benefit of corruption being lower than benefits from holding office (ability to influence policy, public recognition, etc.).[15]

Once again, the only strategic actor is the incumbent, who decides whether to engage in corruption and run for reelection based on the assignment to the high- or low-awareness electorate. The equilibrium concept is the same. The payoff structure is once again left unspecified, as any payoffs which are consistent with the preferences of the incumbent and the voter solve for the same equilibrium of the game, using backward induction.

Figure A3: Second model – corruption and reelection bid are strategic



As in Figure A2, thick lines represent the sub-game perfect Nash equilibrium. Only the incumbent facing the high-awareness voter decides not to engage in corruption, and all incumbents run for reelection on the equilibrium path. The implications for the identification of the results are discussed in the paper.

[14] In principle, the two decisions can be collapsed into one compound decision. I separate the two decisions for the purpose of clarity.

[15] More formally, let b_1 denote the benefits from holding office, and b_2 benefits from engaging in corruption. Define c_1 and c_2 as costs of holding office (for example opportunity costs and time away from family) and of engagement in corruption, respectively. The preferences of the incumbent outlined above are consistent with $b_1 - c_1 > 0, b_2 - c_2 > 0$, and $b_1 > b_2 - c_2$. I believe that these assumptions are uncontroversial.

In both models, it is straightforward to see that the implications of the models are qualitatively the same if H_{A2} – that low-awareness voters are more likely to punish corrupt politicians – is assumed to be true instead of H_{A1}. In this case, the equilibrium play thick lines would be reversed for the two nodes indicating the high- and the low-awareness electorate. For example, in the first model, corrupt incumbents would only choose $\neg R$ when facing a low-information electorate, since choosing otherwise would imply electoral defeat.

References

Bartels, Larry. 1996. "Uninformed Voters: Information Effects in Presidential Elections." *American Journal of Political Science* 40 (1): 194-230.

Bennett, Stephen Earl. 2001. "Reconsidering the Measurement of Political Knowledge Revisited: A Response to Jeffery Mondak." *American Review of Politics* 22: 327-48.

Brown, Lara M. 2006. "Revisiting the Character of Congress: Scandals in the US House of Representatives, 1966-2002." *Journal of Political Marketing* 5 (1): 149–172.

Congressional Quarterly. 1992. *Congressional Ethics: History, Facts, and Controversy*. Washington, D.C.: Congressional Quarterly Press.

Congressional Quarterly Almanac. Various. "Online Editions, `http://library.cqpress.com/cqalmanac/`.".

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, NY: Cambridge University Press.

Hirano, Shigeo, and James M Snyder. 2012. "What Happens to Incumbents in Scandals?" *Quarterly Journal of Political Science* 7: 447–456.

Luskin, Robert C., and John G. Bullock. 2004. "Re(:)Measuring Political Sophistication." Paper presented at the 2004 American Political Science Association Annual Meeting.

Luskin, Robert C., and John G. Bullock. 2011. ""Don't Know" means "Don't Know": DK Responses and the Public's Level of Political Knowledge." *Journal of Politics* 73: 547-557.

Mondak, Jeffery. 1999. "Reconsidering the Measurement of Political Knowledge." *Political Analysis* 8 (1): 57-82.

Mondak, Jeffery. 2001. "Developing Valid Knowledge Scales." *American Journal of Political Science* 45 (1): 224-38.

Mondak, Jeffery, and Belinda Creed Davis. 2001. "Asked and Answered: Knowledge Levels When We Won't Take 'Don't Know' for an Answer." *Political Behavior* 23 (3): 199-224.

Noyer, Mary Ann. 1995. *Catalog of Congressional Ethics Cases, 1976-1992*. Washington, D.C.: Brookings Institution.

Poole, Keith T., and Howard Rosenthal. 2007. *Ideology and Congress*. New Brunswick, NJ: Transaction Publishers.

Roberds, Stephen C. 1997. "Sex, Money, and Deceit: Incumbent Scandals in U.S. House and Senate Election, 1974-1990." Ph.D. diss. Graduate School of the University of Missouri-St. Louis.

Royston, Patrick. 2004. "Multiple Imputation of Missing Values." *The Stata Journal* 4 (3): 227-241.

Royston, Patrick. 2005. "Multiple Imputation of Missing Values: Update of ice." *The Stata Journal* 5 (4): 527-536.

Royston, Patrick. 2007. "Multiple Imputation of Missing Values: Further Update of ICE, with Emphasis on Interval Censoring." *The Stata Journal* 7 (4): 445-464.

Rubin, Donald B. 1987. *Multiple Imputation for Non-Response in Surveys*. New York, NY: John Wiley and Sons.

Rubin, Donald B, Hal S. Stern, and Vasja Vehovar. 1995. "Handling "Don't Know" Survey Responses: The Case of the Slovenian Plebiscite." *Journal of the American Statistical Association* 90 (431): 822-828.

Sekhon, Jasjeet S. N.d. "The Varying Role of Voter Information Across Democratic Societies." Manuscript.

Sturgis, Patrick. 2006. "Measuring Political Knowledge: Gender, Partial Information, and Propensity to Guess." Paper presented at the 2006 American Political Science Asociation Annual Meeting.

Sturgis, Patrick, Nick Allum, and Patten Smith. 2008. "An Experiment on the Measurement of Political Knowledge in Surveys." *Public Opinion Quarterly* 85 (1): 90-102.

van Buuren, Stef, Hendriek C. Boshuizen, and Dick L. Knook. 1999. "Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis." *Statistics in Medicine* 18: 681-694.

Wright, Gerald C. 1993. "Measuring Vote Choice in the National Election Studies, 1952-88." *American Journal of Political Science* 37 (1): 291–316.

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.