Online Appendix for "Uninformed Voters and Corrupt Incumbents"

Contents

A1 Strategic Engagement in Corruption and Selection Bias

Figure A1 shows the extensive-form game tree of the formal model briefly described in the paper. 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.[1] 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.[2] Next, nature assigns the incumbent to a high-awareness electorate (represented by a representative voter) with probability p.[3] 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 incumbent is corrupt or not.[4] 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.

In Figure A1, 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.[5]

[1] This binary awareness space is used for simplicity; the results would be substantively unchanged if the awareness space was less coarse.

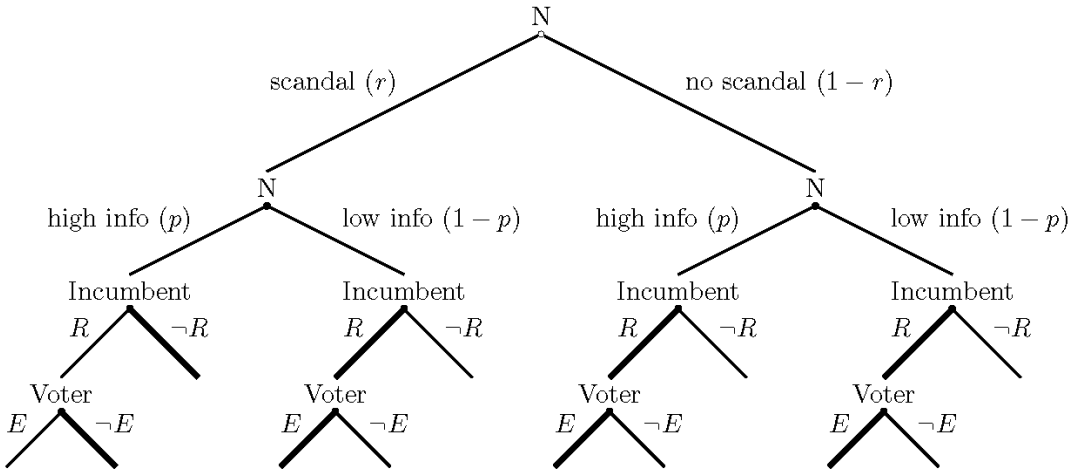
[2] Note that this assignment need not be interpreted as random from the perspective of the voter; it is simply random conditional on political awareness, and the remaining potential factors influencing the decision to engage in corruption are left unmodeled.

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

[4] 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.

[5] 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

Figure A1: First model – reelection bid is strategic



Note: For the assumptions of the model, see the text.

Figure A2 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 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.[6] 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.).[7]

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.

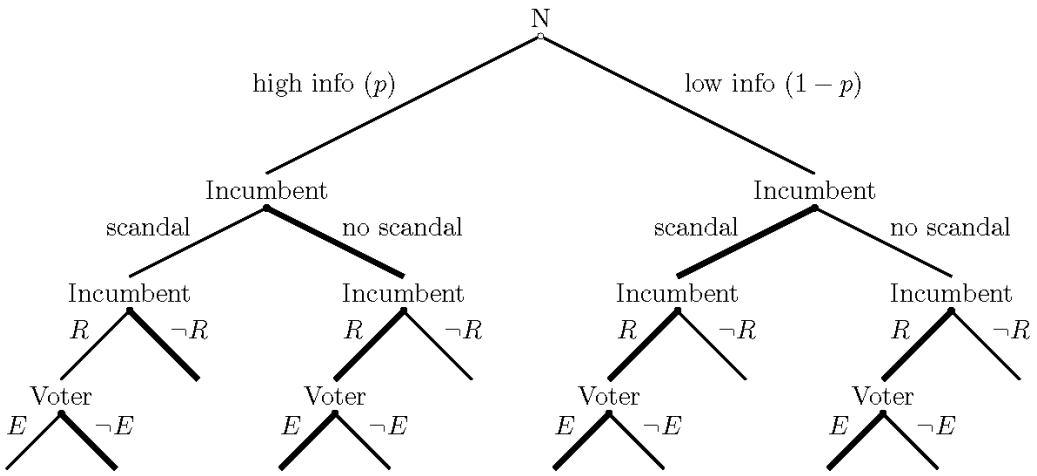
be less likely to run for office in high-awareness districts, precisely where according to H_{A1} voters are more likely to punish corruption.

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

[7] 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.

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 A2: Second model – corruption and reelection bid are strategic



As in Figure A1, 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.

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.

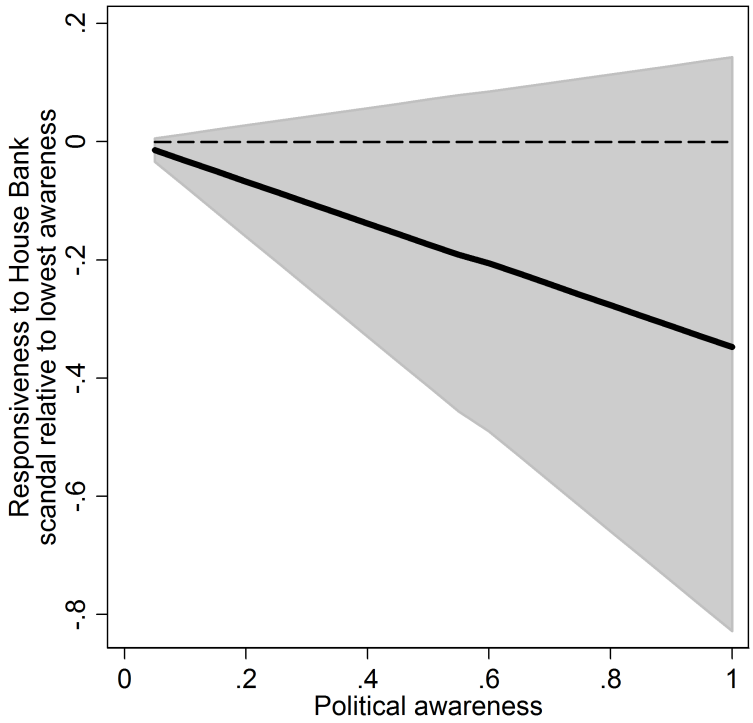
A2 Political Awareness and Corruption Responsiveness

Using the data from the 1992 ANES, Figure A3 shows that after controlling for a rich set of respondent characteristics, low-awareness and high-awareness voters do not statistically significantly differ in the extent to which they consider the House members' behavior in the House Bank scandal punishable. The dependent variable is based on the respondents' opinion about whether the check-writing in the House Bank scandal was punishable.[8] In addition to the independent variables included in the model in equation 1 in the paper, the right-hand side includes a variable denoting

[8]The variable takes the value of: 1 if the respondent thought that: "Writing bad checks is not a serious enough mistake to disqualify someone for office;" 2 if the answer was: "Representatives who wrote only a few bad checks should not be voted out of office just for that reason, but representatives who wrote a lot of bad checks should be voted out of office;" and 3 if the respondent chose: "Representatives who wrote bad checks acted so dishonestly they should be voted out of office."

whether a Congressperson from the respondent's constituency was among the 22 members investigated by the House ethnics committee, as well as the full set of demographic controls,[9] and the interactions between the respondents' income and political awareness. The figure shows the difference in the predicted response between a respondent with a certain level of political awareness, plotted on the x-axis, and a respondent with the lowest level of political awareness (1st percentile). As in the paper, estimates are calculated for values of political awareness from the 1st percentile to the 100th percentile in five-percentile steps. The control variables are kept at the means or medians.

Figure A3: Corruption Responsiveness across Levels of Political awareness



Note: For details, see the text.

A3 Details of Corruption Cases

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

[9]The demographic controls include: age, gender, ethnicity (black and hispanic), income, eduaction, union membership, home ownership, employment status, religion, the census region and the urban status of the place of residence.

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)

Note: 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.

Source: Brown (2006); CQ (Various 1992); Hirano and Snyder (2012); Noyer (1995); Roberds (1997).

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

continued on next page

continued from previous page

Year	Representative	State	District	Description
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.
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.

continued on next page

continued from previous page

Year	Representative	State	District	Description
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.
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.
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.

continued on next page

continued from previous page

Year	Senator	State	Description
1998	Carole Moseley-Braun	IL	FEC investigation on accusation of misuse of $200,000 of leftover campaign funds for personal benefit.

Source: Brown (2006); CQ (Various 1992); Hirano and Snyder (2012); Noyer (1995); Roberds (1997).

A4 Construction of the Political Awareness Score

I construct an individual-level measure of political awareness following the established practice of using knowledge and policy issue questions from surveys (Althaus, 2003; Delli Carpini and Keeter, 1996; Zaller, 1992). Specifically, I rely 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), proper placement of parties and candidates on a left-right scale on policy issues such as health care and defense, and the interviewer's judgement about the respondents' overall political information. 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").[10] The remaining missing values are multiply imputed along with missing values in the other respondent-level variables used in the analysis (see below). 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.[11] 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 broken randomly). 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.[12] Average reliability score, as measured by the Chronbach's α, for the measure is .86.[13]

[10]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).

[11]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.

[12]Normalizing to mean zero and standard deviation of one is another option, but this measure is inappropriate 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.

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

A5 Construct Validity of the Political Awareness Score

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). I use this information to test whether the measure of political awareness is a valid measure that incorporates the familiarity with and understanding of incumbent corruption. To build the appropriate variables, I cross-check the respondents' answers with the data on the actual check-writing by individual legislators.[14] Correct answers receive the value of one, and incorrect the value of zero.[15] I then run a probit model of this binary variable for each item separately on political awareness, 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 below). The results are shown in Figure 1 in the paper.[16]

A6 Multiple Imputation

Since ANES suffers from considerably 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 further 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 feeling thermometer scores (Royston, 2007). Imputations are performed

[14] Overdraft data is taken from "Voters Enraged Over House Bank Abuses." CQ Press Electronic Library, CQ Almanac Online Edition, `http://library.cqpress.com/cqalmanac/cqal92-1106904` (accessed February 28, 2014).

[15] 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. The results are relatively 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.

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

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.

A7 Coding of Covariates and Summary Statistics

Summary statistics for the key variables for the House and for the Senate are given in Tables A4 and A5, respectively.

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

For the individual-level variables, respondents' party identification relative to the incumbent and the self-reported strength of partisanship are combined into one variable, so that strong in-partisans (respondents with the same party ID as the incumbent's party affiliation) receive the value of three, independents the value of zero, strong out-partisans the value of minus three, and weak and leaning in- and out-partisans the remaining values in-between. 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 one if a respondent's economic situation is perceived as "better," zero if "the same," and minus one 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 one, and zero 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 included in the model 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 one if the incumbent is a Democrat, and zero 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.

A8 Coefficient Estimates

Table A6 gives the parameter estimates from the model shown in the paper.

A9 Robustness Checks

Tables A7 and A8 show that the results from Table A6 for the interaction between political awareness and scandal are not sensitive to a number of potential confounds.

Column 1 of each table shows that the results are unaffected if the individual-level characteristics are interacted with the partisanship variable in addition to political awareness, as in the baseline model. Column 2 of the tables indicates that interacting other variables in the model with the corruption indicator does not remove the conditional effect of corruption with respect to political

Table A6: Estimates from the main model

	House		Senate	
Pol. awareness	0.20	(0.21)	0.12	(0.28)
Scandal*aware.	−0.61	(0.26)	−0.41	(0.14)
Scandal	0.13	(0.18)	0.13	(0.13)
Incumbent PID	−0.09	(0.12)	0.13	(0.18)
Incumbent PID*aware.	−0.03	(0.12)	−0.06	(0.13)
Redistricted	−0.12	(0.07)		
PID	0.26	(0.02)	0.24	(0.02)
Econ. eval.	0.04	(0.04)	0.01	(0.06)
Coattails	0.28	(0.08)	0.33	(0.09)
Distance	0.01	(0.02)	0.01	(0.03)
Vote margin	0.10	(0.03)	0.02	(0.04)
Tenure	−0.03	(0.05)	0.05	(0.06)
Campaign intensity	0.01	(0.13)	0.19	(0.26)
Vote margin*aware.	0.04	(0.06)	0.11	(0.06)
Tenure*aware.	−0.02	(0.08)	−0.04	(0.09)
Econ. eval.*aware.	−0.00	(0.07)	−0.02	(0.08)
Coattails*aware.	0.16	(0.12)	0.22	(0.14)
PID*aware.	0.04	(0.03)	0.12	(0.03)
Distanec*aware.	−0.15	(0.03)	−0.11	(0.05)
Intensity*aware.	0.01	(0.21)	−0.48	(0.37)
Constant	0.10	(0.13)	−0.19	(0.20)
Imputations	15		15	
Observations	18996		11060	
ePCP	68.44		65.19	
ePMC	58.60		52.03	
ePRE	23.77		27.43	

Note: The dependent variable is the binary incumbent vote. Standard errors are clustered by district or state. Both models include the full set of year dummies, which are omitted from the table. ePCP is the expected percent correctly predicted (Herron, 2000); ePMC is the expected percent in modal category (Morrison, 1969); ePRE is the expected percent reduction in error, or the difference between the ePCP and ePMC. All measures of fit are averaged across imputations.

awareness, i.e. the moderating effect of political awareness is not confounded by the moderating effect of other variables in the model. Finally, column 3 of the tables shows that political awareness is not simply picking up the influence of respondents' socio-economic status, as proxied by income. Results are further insensitive to the inclusion of a number of demographic characteristics and their interactions with political awareness (results are available upon request).

Differences between low- and high-awareness voters presented in the paper may potentially be explained not by differences in political awareness but because constituencies with different levels of political awareness face incumbents engaged in different types of corruption due to incumbent

Table A7: Adding potential confounds – House

	(1)		(2)		(3)	
Pol. awareness	0.20	(0.21)	0.19	(0.21)	0.23	(0.25)
Scandal*aware.	−0.62	(0.26)	−0.61	(0.26)	−0.58	(0.26)
Scandal	0.14	(0.18)	0.25	(0.31)	0.27	(0.35)
Econ. eval.*PID	0.00	(0.02)	0.00	(0.02)	0.00	(0.02)
Coattails*PID	−0.06	(0.03)	−0.06	(0.03)	−0.06	(0.03)
Distance*PID	−0.00	(0.01)	−0.00	(0.01)	−0.00	(0.01)
Econ. eval.*PID*aware.	−0.01	(0.03)	−0.00	(0.03)	−0.00	(0.03)
Coattails*PID*aware.	0.09	(0.06)	0.09	(0.06)	0.09	(0.06)
Distance*PID*aware.	0.02	(0.02)	0.02	(0.02)	0.02	(0.02)
PID*scandal			−0.04	(0.03)	−0.04	(0.03)
Econ. eval.*scandal			−0.03	(0.07)	−0.02	(0.07)
Coattails*scandal			0.03	(0.12)	0.03	(0.12)
Distance*scandal			0.03	(0.04)	0.03	(0.04)
Vote margin*scandal			−0.04	(0.08)	−0.05	(0.08)
Tenure*scandal			−0.02	(0.09)	−0.01	(0.09)
Intensity*scandal			−0.32	(0.29)	−0.34	(0.29)
Income, 17-33 pct					0.06	(0.10)
Income, 34-67 pct					−0.03	(0.08)
Income, 68-95 pct					0.02	(0.09)
Income, 96-100 pct					0.11	(0.16)
Income, 17-33 pct*aware.					−0.10	(0.19)
Income, 34-67 pct*aware.					−0.01	(0.15)
Income, 68-95 pct*aware.					−0.01	(0.16)
Income, 96-100 pct*aware.					−0.20	(0.23)
Income, 17-33 pct*scandal					0.06	(0.19)
Income, 34-67 pct*scandal					0.06	(0.18)
Income, 68-95 pct*scandal					−0.15	(0.17)
Income, 96-100 pct*scandal					0.09	(0.27)
Constant	0.11	(0.13)	0.11	(0.13)	0.10	(0.15)
Imputations	15		15		15	
Observations	18996		18996		18996	
ePCP	68.51		68.53		68.54	
ePMC	58.60		58.60		58.60	
ePRE	23.94		23.97		24.01	

Note: The dependent variable is the binary incumbent vote. Standard errors are clustered by district. All models include the original control variables indicated in the main text, but are omitted from the table to save space. ePCP is the expected percent correctly predicted (Herron, 2000); ePMC is the expected percent in modal category (Morrison, 1969); ePRE is the expected percent reduction in error, or the difference between the ePCP and ePMC. All measures of fit are averaged across fifteen imputations.

Table A8: Adding potential confounds – Senate

	(1)		(2)		(3)	
Pol. awareness	0.12	(0.28)	0.11	(0.28)	0.10	(0.33)
Scandal*aware.	−0.41	(0.14)	−0.38	(0.14)	−0.37	(0.14)
Scandal	0.14	(0.13)	0.98	(0.38)	0.98	(0.41)
Econ. eval.*PID	−0.00	(0.02)	−0.00	(0.02)	−0.00	(0.02)
Coattails*PID	−0.05	(0.04)	−0.05	(0.04)	−0.04	(0.04)
Distance*PID	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)
Econ. eval.*PID*aware.	−0.03	(0.04)	−0.03	(0.04)	−0.03	(0.04)
Coattails*PID*aware.	0.04	(0.07)	0.04	(0.07)	0.03	(0.07)
Distance*PID*aware.	−0.01	(0.02)	−0.01	(0.02)	−0.01	(0.02)
PID*scandal			0.01	(0.04)	0.01	(0.04)
Econ. eval.*scandal			−0.13	(0.07)	−0.13	(0.07)
Coattails*scandal			−0.08	(0.12)	−0.08	(0.11)
Distance*scandal			0.03	(0.06)	0.03	(0.06)
Vote margin*scandal			−0.09	(0.09)	−0.09	(0.09)
Tenure*scandal			−0.30	(0.10)	−0.30	(0.10)
Intensity*scandal			−0.41	(0.70)	−0.44	(0.69)
Income, 17-33 pct					0.11	(0.14)
Income, 34-67 pct					0.05	(0.12)
Income, 68-95 pct					0.11	(0.14)
Income, 96-100 pct					0.11	(0.21)
Income, 17-33 pct*aware.					−0.09	(0.23)
Income, 34-67 pct*aware.					0.03	(0.22)
Income, 68-95 pct*aware.					−0.07	(0.26)
Income, 96-100 pct*aware.					0.13	(0.31)
Income, 17-33 pct*scandal					−0.03	(0.23)
Income, 34-67 pct*scandal					0.01	(0.18)
Income, 68-95 pct*scandal					0.01	(0.18)
Income, 96-100 pct*scandal					−0.17	(0.29)
Constant	−0.18	(0.20)	−0.25	(0.21)	−0.31	(0.22)
Imputations	15		15		15	
Observations	11060		11060		11060	
ePCP	65.24		65.35		65.39	
ePMC	52.03		52.03		52.03	
ePRE	27.53		27.76		27.84	

Note: The dependent variable is the binary incumbent vote. Standard errors are clustered by state. All models include the original control variables indicated in the main text, but are omitted from the table to save space. ePCP is the expected percent correctly predicted (Herron, 2000); ePMC is the expected percent in modal category (Morrison, 1969); ePRE is the expected percent reduction in error, or the difference between the ePCP and ePMC. All measures of fit are averaged across fifteen imputations.

characteristics omitted from the empirical model. This concern can be addressed with the inclusion of candidate fixed effects. Since no politician faces strictly a fully informed or uninformed electorate, the interaction between corruption and political awareness is still identified in a model that includes candidate fixed effects as long as there are observations in the ANES from multiple elections under the same candidate. The estimation sample in this case is restricted to districts or states that had a corrupt incumbent at some point, and the analysis requires data on voters both before and after the charge of corruption. Because the ANES uses clustered samples, many congressional districts and states indeed have repeated observations. For the House of Representatives, the 97 unique congressional districts covered in the ANES that had at least one corrupt incumbent over the study period (for a total of 107 corruption cases – see Table A3 above) had 263 distinct incumbents represented by at least one respondent in the ANES. Of these, 193 candidates are represented by at least one respondent in repeated surveys in the ANES, giving approximately 6,200 observations for estimation. For the Senate, the 14 states covered in the ANES that had at least one corrupt incumbent over the study period (for a total of 17 corruption cases) had 67 distinct incumbents represented by at least one ANES respondent. Of these 37 incumbents are represented by respondents in repeated surveys, giving approximately 5,200 observations for estimation. The results with these samples are substantively unchanged, and are available upon request.

Another way to address the concern of candidate heterogeneity due to unobserved factors correlated with corruption is to conduct a placebo test by comparing high- and low-awareness voters in an election before the charge of corruption. If political awareness, rather than some unobservable characteristic moderates the effect of corruption, we should not observe systematic differences between voters with different levels of political awareness in the absence of corruption. As with the model with candidate fixed effects, this test requires repeated observations in the same districts with corrupt incumbents. For the House of Representatives, there are 923 respondents that fit the requirements of the test – respondents in districts in election years immediately prior to a formal corruption charge. For the Senate, the sample size is 208 respondents. The regressions now exclude the scandal variable and the interaction, and the relevant test is simply the coefficient on political awareness. This placebo test shows no statistically significant effect of political awareness, suggesting that it is indeed corruption that differently affects voters with different levels of political awareness. This finding also lends further support – in addition to the evidence in Figure 1 in the paper – to the claim that political awareness is strongly correlated with familiarity with incumbent corruption. It suggests that low-awareness voters fail to punish corruption because they are less aware of it, rather than because they act differently than high-awareness voters even when they are aware of corruption.

The baseline model also omits potentially important aggregate-level confounders such as challenger quality and campaign spending. Because of the long period of coverage, consistent measures of these characteristics are unavailable.[17] Including district or state fixed effects may absorb some of these unmeasured factors, even if in an imperfect way.[18] The results are substantively unchanged, and are available upon request. The more restricted specification in the baseline model is used

[17]Restricting the period of coverage so that such measures are available is not possible because too many observations with corrupt incumbents are discarded.

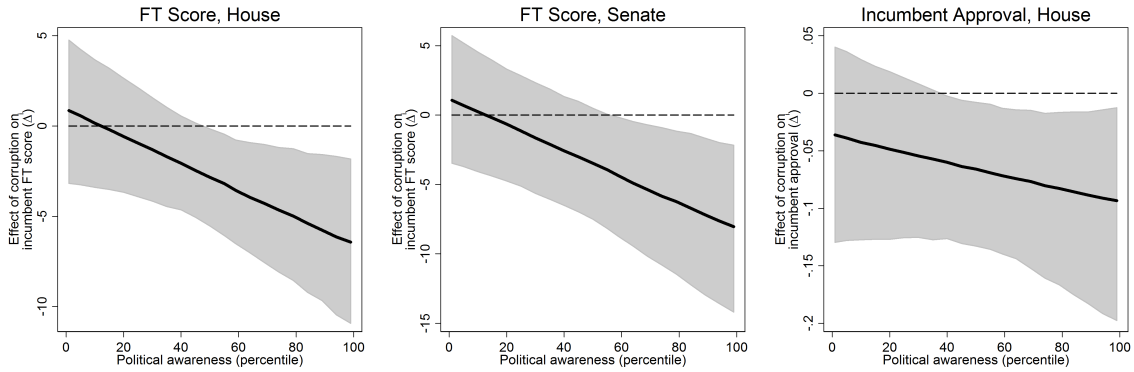
[18]Since fixed effects are perfectly collinear with the aggregate-level controls from the baseline model, I exclude the latter. Moreover, because of a small number of individuals in each district or state in ANES, including fixed effects induced a number of perfect predictions, whereby the parameter estimates are set to negative infinity. Stata drops these observations and variables that induce perfect prediction of the outcome. To avoid this, I augment the data with a simple ad-hoc procedure proposed by White, Daniel and Royston (2010).

primarily due to high computational costs of estimating a model with a large number of fixed effect parameters.

It is difficult to fully rule out the omitted variable bias using observational data. However, an approach described in Altonji, Elder and Taber (2005) allows for a calculation of how much greater the influence of unobservable variables would need to be, relative to observed variables, to completely explain away the identified negative association between awareness and corruption and incumbent support. The idea is to measure how the main coefficients of interest change following the inclusion of observed controls, and use this change as the proxy for the effect of the unobserved variables, had they been included in the model.[19] I find that such omitted variable bias would have to be 2.6 and 2.9 times as large as the effect of included observable confounders for the House and the Senate, respectively, to explain away the entire composite effect of corruption and political awareness on incumbent support. Since the common predictors of incumbent support are already included, this magnitude of omitted variable bias seems unlikely.

Finally, the dependent variable in the paper 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.

Figure A4: Alternative measures of incumbent support



Note: The line represents the estimates of the difference in support for a corrupt incumbent relative to a clean incumbent for different levels of political awareness. The model is identical to the baseline model in equation 1 in the paper, expect that the stated vote choice is replaced with the incumbent feeling thermometer scores (left and center panels), or incumbent approval (the right-most panel). When the feeling thermometer is the dependent variable, the models are OLS regressions rather than probit models as in the baseline case. All other variables in the model are kept at their means or medians. Negative values indicate a negative impact of corruption on incumbent support for 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 95 percent confidence interval. All quantities are averaged over fifteen imputations.

As alternative measures, I use incumbent approval (only available for the House), and incumbent feeling thermometer scores. The former is a standard binary measure; the latter lies in the $[0, 100]$

[19] Details on the calculation of this summary measure, the underlying assumptions, and the proof of the application of this concept to the specification with interaction terms are available upon request. It will be posted on the author's website when the paper is published.

interval.[20] Figure A4 shows the same individual effect results as in Figure 1 in the paper, but with stated vote choice replaced with these alternative measures of incumbent support.[21] The results are quite similar to those for incumbent vote. An FT score given to a corrupt incumbent compared to a clean incumbent decreases as political awareness of an average voter increases. This results holds both for the House and the Senate. Similarly, a high-awareness voter is less likely to approve of a corrupt representative in the House than a low-awareness voter.

The remaining robustness checks are outlined in the main text. All results are available upon request.

References

Althaus, Scott L. 2003. *Collective Preferences in Democratic Politics: Opinion Surveys and the Will of the People*. New York: Cambridge University Press.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *The Journal of Political Economy* 113(1):151–184.

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.

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

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

Delli Carpini, Michael X. and Scott Keeter. 1996. *What Americans Know about Politics and Why It Matters?* New Heaven: Yale University Press.

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

Herron, Michael C. 2000. "Postestimation Uncertainty in Limited Dependent Variable Models." *Political Analysis* 8(1):83–98.

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

[20] 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.

[21] For the FT score variable, OLS is fitted instead of the probit model.

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.

Morrison, Donald G. 1969. "On the Interpretation of Discriminant Analysis." *Journal of Marketing Research* 6(2):156–163.

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 PhD thesis 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.

White, Ian R., Rhian Daniel and Patrick Royston. 2010. "Avoiding Bias due to Perfect Prediction in Multiple Imputation of Incomplete Categorical Variables." *Computational Statistics and Data Analysis* 54:2267–2275.

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.