The Age Penalty - Evidence from Two Survey Experiments

George Quinn[†*]

[†]PhD Candidate. Rutgers University. Department of Political Science.

george.quinn@rutgers.edu

*

Abstract

Candidate age is prominent in American politics, yet its independent effect is difficult to separate from experience, incumbency, and correlated attributes. This paper reports two survey experiments involving 4,128 U.S. adults who evaluated congressional or gubernatorial candidates assigned ages from 25 to 85. Age produced the largest and only statistically detectable demographic effect in Study 1, but the corresponding linear estimate in Study 2 was negative and imprecise. Across both studies, however, evaluations were stable between ages 30 and 50 and declined at ages 70 and 80. An 80-year-old candidate was rated 0.26 standard deviations lower than a 50-year-old in Study 1 and 0.14 lower in Study 2. Insider–outsider biographies and respondents' own ages did not detectably moderate this pattern. Open-ended responses linked higher assigned ages to age references and concerns about health, adaptability, and capacity, revealing a nonlinear advanced-age penalty rather than a broadly uniform decline across adulthood.

1 Introduction

Candidate age has become increasingly salient in American politics. At the beginning of the 118th Congress, members of the U.S. House averaged 57.9 years of age and senators averaged 64.0 years (Manning, 2024). In 2023, 79% of Americans supported maximum age limits for federal elected officials, with majorities in both parties favoring the proposal (Gramlich, 2023). Yet the political visibility of age does not establish whether a candidate's stated numerical age independently affects voter evaluations. Observational comparisons are difficult because age is ordinarily bundled with political experience, incumbency, name recognition, career duration, and institutional resources. The age composition of elected office may also reflect candidate emergence, party recruitment, career trajectories, and younger citizens' willingness to seek office (Lawless and Fox, 2015; Shames, 2017; Stockemer and Sundström, 2023). These processes help explain who reaches the ballot, but they are distinct from the demand-side question examined here: once voters encounter a candidate, does stated age itself change how they evaluate that candidate?

Age is a multidimensional political cue. Younger candidates may be associated with energy, innovation, and generational change, but they may also be perceived as insufficiently prepared. Older candidates may benefit from associations with experience, accumulated knowledge, and judgment, while candidates at more advanced ages may prompt concerns about health, stamina, adaptability, cognitive capacity, or the ability to complete the responsibilities of office. Voters commonly use demographic and biographical characteristics to infer qualities that are not directly observed (Popkin, 1991; Lau and Redlawsk, 2001; Kirkland and Coppock, 2018; McDonald, Karol, and Mason, 2020), and prior research shows that candidate age can shape political judgments (L. Sigelman and C. K. Sigelman, 1982; Riggle and Johnson, 1996; Woods, Tadlock, and Riggle, 2003; Roberts and Wolak, 2023; McClean and Ono, 2024). What remains less clear is whether age has a general linear effect on candidate evaluations particularly at the upper end of the age range.

This article addresses these questions using two survey experiments[1]. Respondents evaluated hypothetical congressional or gubernatorial candidates whose stated ages were randomly assigned across the 25–85 range, alongside an insider outsider biography and other candidate characteristics. Consistent with the first preregistered expectation, age produced the largest and only statistically detectable effect among the randomized demographic attributes in Study 1. In Study 2, however, the corresponding linear age estimate was negative but statistically imprecise, providing only partial support for a general linear age effect. The

[1]Study 1 was pre-registered and fielded in Spring 2024 including 1,909 valid cases, while Study 2 was not pre-registered and therefore conclusions should be considered confirmatory

more consistent finding was that positive evaluations were broadly similar between ages 30 and 50 but declined at ages 70 and 80. This paper finds no detectable evidence that this upper-age decline was moderated by the insider outsider biography or by respondents' own ages. Open-ended responses further show that higher assigned ages increased explicit references to age and concerns associated with advanced age. The findings therefore identify a modest evaluative penalty concentrated at the upper end of the age range rather than a uniform decline across adulthood.

Study 1 was pre-registered in April 2024. The pre-registration document is available at `https://aspredicted.org/djs6-qjvn.pdf`. See *Pre-Registration* within the Appendix for information about the correspondence between the manuscript and the pre-registration document.

2 Candidate Age as an Evaluative Cue

2.1 The competing meanings of candidate age

Voters use demographic and biographical cues to infer candidate qualities that are not directly observed, particularly in low-information settings (Popkin, 1991; Lau and Redlawsk, 2001; Kirkland and Coppock, 2018; McDonald, Karol, and Mason, 2020). These inferences need not be irrational, but they can apply generalized expectations to an individual candidate. Age is an especially multidimensional cue because it can communicate both political assets and liabilities.

Older candidates may be associated with experience, accumulated knowledge, and judgment, but also with reduced adaptability or physical capacity (Fiske et al., 2002; Cuddy, Norton, and Fiske, 2005; North and Fiske, 2012). Younger candidates may raise concerns about preparation while also conveying energy, innovation, and generational renewal. At advanced ages, voters may infer declining stamina, health risks, reduced flexibility, or a shorter prospective horizon in office. Numerical age can therefore activate competing considerations whose net influence may change across the life course.

The observable implications can be represented as

$$\mathrm{E}[Y_i \mid A_i = a, B_i = b] = \alpha_b + f(a) + g_b(a),$$

where Y_i is the candidate evaluation, A_i is the randomly assigned candidate age, and B_i identifies the assigned party-by-biography profile. The term α_b captures baseline differences across profiles, $f(a)$ represents the common net age-response function, and $g_b(a)$ represents any profile-specific departure from that function. This formulation maps directly onto the

empirical analysis: profile means estimate differences in α_b, the pooled age models estimate $f(a)$, and the age-by-profile interactions test whether $g_b(a)$ differs from zero.

The independent variation in biography is crucial to interpreting the age effect. By presenting the candidate as either as an experienced officeholder or as a political newcomer, the experiments distinguish numerical age from the candidate's stated route to office. If voters use age primarily as a proxy for political experience, the age response should differ across those biographies because experience has already been supplied explicitly. If experience instead yields diminishing returns while concerns associated with advanced age continue to grow, evaluations should change little from young adulthood through middle age and then decline at older ages. A similar upper-age decline across the different biographies would therefore identify an advanced-age penalty that cannot be attributed solely to voters using age as a substitute for information about political experience.

2.2 Prior evidence and contribution

Direct research establishes that age can enter candidate judgments, although its effect depends on the surrounding information. Early experiments showed that age can affect evaluation alongside race and gender (L. Sigelman and C. K. Sigelman, 1982). Candidate age also influences the strategies citizens use to evaluate politicians, with possible variation by the evaluator's own age (Riggle and Johnson, 1996). Individualized candidate information can attenuate or alter age-based judgments (Woods, Tadlock, and Riggle, 2003), and more recent work shows that citizens care about the age of elected representatives within particular partisan and electoral contexts (Roberts and Wolak, 2023). Recent experimental evidence further suggests that voters may evaluate younger and middle-aged candidates similarly while penalizing older candidates. McClean and Ono (2024) provide recent experimental evidence on visual-age and show that older voters do not necessarily favor older candidates.

This literature does not imply a universal ideal age or a mechanically symmetric preference for middle age. It instead indicates that age is an evaluative cue whose meaning depends on the other information voters receive. The present design extends this research by randomizing ages continuously from 25 to 85, independently varying formal political experience, repeating the test in congressional and gubernatorial settings, and analyzing respondents' explanations as post-treatment text outcomes.

Below are the seven experimental hypotheses tested in this analysis. Importantly, a hypothesis is noted as preregistered or exploratory depending on the correct documentation.[2]

[2]Wave 1 preregistered the randomized age impacts, biographical effects, and cognitive recall outcomes; it did not preregister respondent-age moderation, or open-ended text analyzes; therefore, Hypotheses 1-3 are preregistered in Wave 1. Wave 2 assesses the same directional patterns without pre-registration.

H1: Age Impacts When testing other candidate demographics, Age will have likely the highest impact on candidate likeness. This is in comparison with varied age, race, and gender. (Pre-Registered)

H2: Biographical Effects When considering the effects of framing, this study measures outsider/insider biographies and seeks to identify whether an outsider/insider dynamic is more profound then just strict candidate age. (Pre-Registered)

H3: Cognitive recall and perceived effectiveness When respondents are asked post exposure what they liked about this candidate, age will be mentioned more then other attributes. In addition when asked to recall these attributes age will be dominant. Post exposure questions include cognitive recall for all demographic factors, positing that respondents will recall age over other attributes. (Pre-Registered)

H4: Asymmetric age response The evaluative decline between ages 50 and 80 will exceed the evaluative difference between ages 30 and 50. (Exploratory)

H5: Respondent-age moderation Candidate age may be interpreted differently across a respondents' life stages. Age similarity could produce perceived commonality, while older respondents may place greater value on experience or interpret advanced-age capacity differently. (Exploratory)

2.3 Open-ended explanations

Closed-ended evaluations reveal the net effect of the competing considerations in the above hypotheses, but not which considerations respondents express. The open-ended responses provide secondary evidence about how the age treatment changes the content of candidate evaluation. Because respondents answered after reporting their closed-ended rating, these responses are treated as post-treatment textual outcomes rather than as measures of mediation.

H6: Age salience As assigned age enters the upper portion of the experimental range, respondents will become more likely to mention the candidate's age explicitly. (Exploratory)

H7: Advanced-age concerns As assigned age increases, respondents will become more likely to express concerns about health, stamina, retirement, adaptability, capacity, or being out of touch, particularly at ages 70 and 80 relative to age 50. (Exploratory)

Each expectation is evaluated separately in the two studies. Pooled wave-by-age interactions assess whether the estimated response functions differ reliably across experiments, but cannot identify whether any difference is caused by year, office, or outcome format.

3 Experimental Design

This study analyzes two survey experiments fielded one year apart. Study 1, conducted in spring 2024, presented respondents with a hypothetical candidate for the U.S. House and measured candidate evaluation on a five-category like–dislike scale. Study 2, conducted in spring 2025, presented a hypothetical gubernatorial candidate and used a 0–100 feeling thermometer. Because the studies differ simultaneously in year, office, and outcome format, Study 2 serves as a replication and generalizability test rather than as a causal comparison between congressional and gubernatorial elections.

Study 1 was administered to U.S. adults through Cint/Lucid from April 10–18, 2024. Of 2,907 imported responses, 2,737 respondents consented and completed the survey, 2,730 passed the attention check, and 2,251 passed the honesty item. The candidate-age experiment included 1,912 respondents with an assigned age between 25 and 85; 1,909 had exactly one usable candidate evaluation and complete treatment information. The 339 post-honesty cases without an assigned candidate age or condition outcome reflect missing experimental exposure rather than exclusion based on treatment response. Study 2 was administered through Cint/Lucid from April 25–May 31, 2025. The final analytical sample contains 2,219 respondents who received exactly one candidate profile, reported a valid 0–100 evaluation, and had complete treatment information.

3.1 Partisan routing and candidate profiles

The experiments placed respondents in a same-party candidate-selection environment. Democrats evaluated a Democratic candidate and Republicans evaluated a Republican candidate. Independents were routed according to their partisan leaning; those who did not initially express a leaning were routed using a forced-choice party item. Every respondent with a resolved party preference was matched to the corresponding partisan profile in both studies.[3]

Within these routed partisan environments, respondents were randomly assigned to an insider or outsider biography. The insider was described as a state senator who had served multiple terms and understood the legislative process. The outsider was described as a

[3]The Appendix provides the complete treatment text for every party-by-experience profile.

successful entrepreneur and political newcomer offering a fresh, results-oriented perspective. The profiles also independently randomized Alex Taylor's stated numerical age from 25 to 85, gender, race (White, Black, Hispanic, or Asian), and ideological position. Democratic candidates were described as progressive or moderate, whereas Republican candidates were described as conservative or moderate. Study 1 described Taylor as running for Congress; Study 2 described Taylor as running for governor.

The final analytical samples include 1,909 respondents in Study 1 and 2,219 in Study 2. The four party-by-experience profiles were approximately balanced within each study, and respondents averaged about 50 years of age in both samples. The latter permits an exploratory assessment of whether candidate-age evaluations vary with respondent age.

3.2 Outcomes and estimation

The primary outcome in Study 1 ranges from 1 ("dislike a great deal") to 5 ("like a great deal"). Study 2 uses a 0–100 feeling thermometer. Results are reported in each study's original units. For pooled comparisons, the outcomes are placed on a common 0–1 scale: Study 1 is transformed as $(Y_i - 1)/4$, and Study 2 as $Y_i/100$. Within-study standardized outcomes are used to express comparable effect sizes.

After reporting the closed-ended evaluation, respondents provided an open-ended explanation of their impression of the candidate. These responses are analyzed as randomized post-treatment text outcomes. Study 1 also included a separate measure of the candidate's perceived effectiveness in office.

The primary specification estimates

$$Y_i = \alpha + \beta_1 \widetilde{Age}_i + \beta_2 \widetilde{Age}_i^2 + \boldsymbol{\gamma}'\mathbf{X}_i + \varepsilon_i,$$

where $\widetilde{Age}_i = (Age_i - 55)/10$. The vector $\mathbf{X}_i$ includes the randomized experience, ideology, race, and gender indicators, along with the respondent's routed partisan environment. Models are estimated using ordinary least squares with HC2 standard errors.

The coefficients are translated into average predicted evaluations and contrasts at ages 30, 50, 70, and 80. These comparisons operationalize the distinction between young adulthood, middle age, and the upper portion of the experimentally assigned range without imposing a categorical threshold for "advanced age." Natural-spline specifications assess whether the substantive conclusions depend on the quadratic approximation. Because year, office, and outcome format all change between studies, this interaction tests the consistency of the estimated age-response functions rather than the causal effect of electoral office.

4 Results

4.1 H1: Age Impact Models

The first analysis compares the effect of assigned candidate age with the effects of the other randomized candidate attributes. The dependent variable is standardized separately within each study, as is assigned candidate age. The linear specifications estimate the change in candidate evaluation associated with a one-standard-deviation increase in age. The quadratic specifications add squared standardized age to allow the response to vary over the experimental range. All models use OLS with HC2-robust standard errors. Moderate ideology, a White candidate, and the male pronoun are the reference categories. Table 1 summarizes the coefficient estimates[4].

Table 1: Randomized candidate attributes and standardized candidate evaluations

	Study 1: Spring 2024		Study 2: Spring 2025	
Randomized attribute	Linear model	Quadratic model	Linear model	Quadratic model
Intercept	−0.031	0.047	−0.045	−0.011
	(0.055)	(0.059)	(0.053)	(0.058)
Candidate age (1 SD)	−0.056*	−0.056*	−0.035	−0.033
	(0.024)	(0.024)	(0.021)	(0.021)
Candidate age squared	—	−0.077**	—	−0.035
		(0.026)		(0.025)
Progressive/conservative ideology	0.066	0.068	−0.028	−0.026
	(0.046)	(0.046)	(0.042)	(0.042)
Black candidate	−0.032	−0.034	0.148*	0.149*
	(0.066)	(0.065)	(0.061)	(0.061)
Hispanic candidate	0.000	−0.004	0.101	0.103
	(0.061)	(0.061)	(0.061)	(0.061)
Asian candidate	−0.034	−0.036	0.133*	0.135*
	(0.064)	(0.064)	(0.060)	(0.060)
Female pronoun	0.028	0.028	−0.073	−0.074
	(0.046)	(0.046)	(0.042)	(0.042)
Observations	1,909	1,909	2,219	2,219

Notes: Entries are OLS coefficients with HC2-robust standard errors in parentheses. Candidate evaluations and assigned candidate age are standardized separately within each study. Race coefficients are relative to a White candidate; ideology is relative to a moderate candidate; and the female-pronoun coefficient is relative to the male pronoun. Study 1 is the preregistered experiment, and Study 2 is the confirmatory replication. $^{*}p < .05$; $^{**}p < .01$.

The Study 1 results provide the clearest coefficient-level evidence that assigned age affected candidate approval. In the linear model, a one-standard-deviation increase in candidate age reduced the evaluation by 0.056 standard deviations (HC2 SE = 0.024, 95% CI $[-0.103, -0.008]$, $p = .0209$). None of the ideology, race, or pronoun coefficients reached conventional significance. The linear model's joint test of all randomized attributes did not

[4]Empirical tests for this hypothesis were not pre-registered, ATE analysis was initially planned but did not fit the model for best analysis. Full model results are located in Appendix C.1

reject the null, however, $\chi^2(6) = 8.41$, $p = .209$, and the model explained little of the total respondent-level variation ($R^2 = .0047$).

Allowing age to enter quadratically strengthens the evidence that the age response is nonlinear. The Study 1 linear age component remains negative ($b = -0.056$, HC2 SE $= 0.024$, $p = .0195$), while the squared age term is also negative and statistically significant ($b = -0.077$, HC2 SE $= 0.026$, 95% CI $[-0.128, -0.027]$, $p = .0028$). Study 2 reproduces the negative direction of the age coefficients but not their individual statistical significance. The linear model estimates a 0.035-standard-deviation decline for a one-standard-deviation increase in age (HC2 SE $= 0.021$, 95% CI $[-0.077, 0.007]$, $p = .102$). In the quadratic model, neither the linear age component ($b = -0.033$, $p = .121$) nor the squared term ($b = -0.035$, $p = .152$) is individually distinguishable from zero. By contrast, Black and Asian candidate profiles receive higher evaluations than otherwise equivalent White profiles in both Study 2 specifications. In the linear model, the differences are 0.148 standard deviations for Black candidates (95% CI $[0.029, 0.267]$, $p = .0147$) and 0.133 for Asian candidates (95% CI $[0.016, 0.251]$, $p = .0263$). These estimates remain nearly unchanged after the squared age term is added. The Study 2 specifications are jointly significant (linear: $\chi^2(6) = 13.33$, $p = .038$; quadratic: $\chi^2(7) = 15.09$, $p = .035$), although their explanatory power remains small ($R^2 = .0062$ and $.0072$, respectively).

These candidate-approval models provide partial support for H1. Age is the only randomized candidate attribute with a statistically detectable coefficient in Study 1, and the quadratic specification identifies significant negative curvature. Study 2 produces age coefficients in the same negative direction, but race rather than age is the more prominent coefficient-level predictor. Section 4.4 therefore evaluates the age response through predicted values and selected-age contrasts, which show where along the age range the evaluative penalty is concentrated.

4.2 H2: Biographical Effects

The insider–outsider manipulation addresses two distinct questions: whether experience changes candidates' average evaluations and whether it changes how voters respond to candidate age. Experience could affect baseline ratings without either buffering or intensifying an advanced-age penalty. Figure 1 reports the unadjusted means.

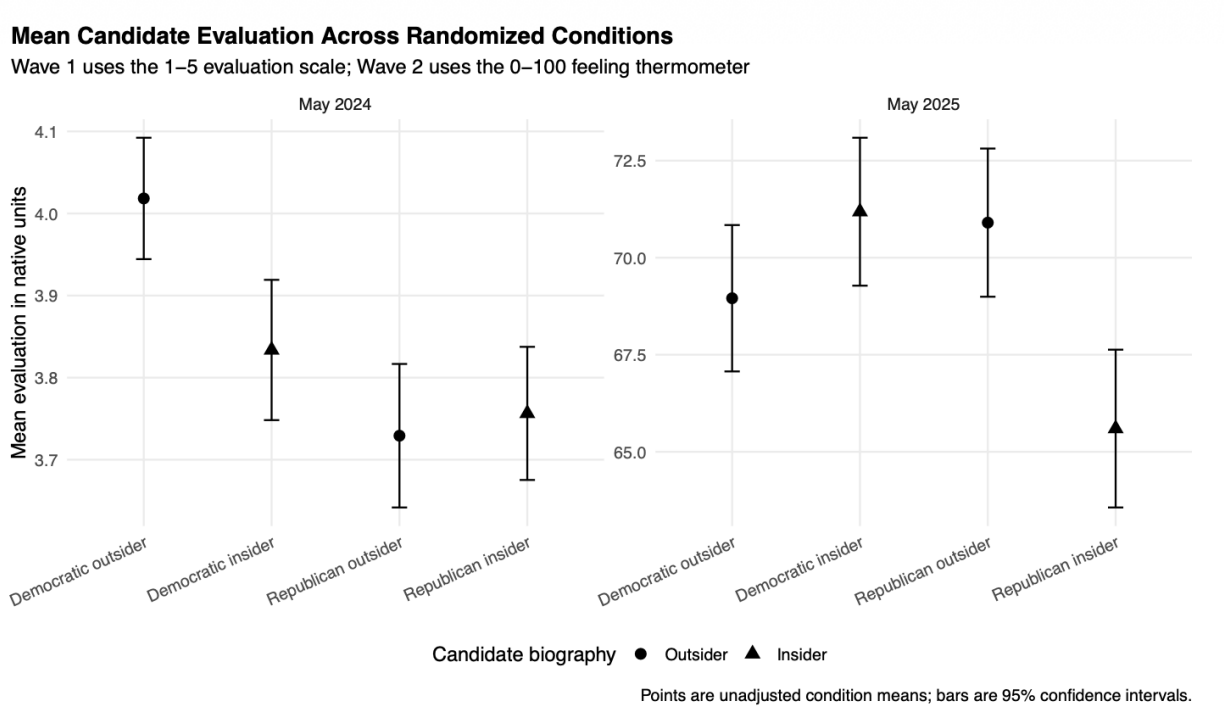

Figure 1: Mean candidate evaluations across randomized party-by-biography profiles. Study 1 reports evaluations on the original 1–5 scale; Study 2 reports evaluations on the original 0–100 feeling thermometer. Points are unadjusted condition means, and bars are 95% confidence intervals. Circles identify outsider profiles and triangles identify insider profiles.

In Study 1, the Democratic insider was evaluated 0.185 points less favorably than the Democratic outsider on the 1–5 scale (95% CI: $-0.297, -0.072$; Holm-adjusted $p = .0027$), equivalent to 4.62 points on a 0–100 scale. The Republican contrast was small and non-significant ($\Delta = 0.027$, $p = .654$). In Study 2, the Republican insider was evaluated 5.305 points less favorably than the Republican outsider (95% CI: $-8.087, -2.523$; Holm-adjusted $p = .0004$), whereas the Democratic contrast was not significant ($\Delta = 2.229$, $p = .102$). Thus, detectable differences favored outsiders, but they appeared in different parties across studies; the results do not indicate a uniform insider or outsider advantage.

The second question is whether experience changes the age response. Models allowed both the linear and quadratic age terms to vary across the four profiles, and a six-degree-of-freedom joint test evaluated the resulting interactions. The age-response functions did not differ significantly in Study 1, $\chi^2(6) = 2.76$, $p = .838$, or Study 2, $\chi^2(6) = 8.45$, $p = .207$. No condition-specific age-function test survived adjustment for multiple comparisons. The joint interaction is the relevant moderation test: a significant age coefficient in one subgroup and a nonsignificant coefficient in another would not itself show that the subgroups differ.

Table 2: Experience-profile contrasts and tests of age-response heterogeneity

Study	Test or contrast	Estimate or statistic	p
Spring 2024	Omnibus difference across four profile means	$\chi^2(3) = 32.64$	$< .001$
Spring 2024	Democratic insider minus outsider	-0.185 $[-0.297, -0.072]$	.0027
Spring 2024	Republican insider minus outsider	$+0.027$ $[-0.092, 0.146]$	.6540
Spring 2024	Equality of four age-response functions	$\chi^2(6) = 2.76$	.8381
Spring 2025	Omnibus difference across four profile means	$\chi^2(3) = 19.42$	.0002
Spring 2025	Democratic insider minus outsider	$+2.229$ $[-0.446, 4.904]$	.1024
Spring 2025	Republican insider minus outsider	-5.305 $[-8.087, -2.523]$	.0004
Spring 2025	Equality of four age-response functions	$\chi^2(6) = 8.45$	.2072

Notes: Insider-minus-outsider differences are reported in each study's native outcome units, with 95% confidence intervals in brackets. Study 1 uses the 1–5 candidate-evaluation scale; Study 2 uses the 0–100 feeling thermometer.

Experience profiles therefore shifted some baseline evaluations but neither reliably shielded older candidates from the age penalty nor made that penalty more severe. Given the absence of moderation effects, biographical variation does not explain differences in candidate evaluations.

4.3 H3: Cognitive recall and perceived effectiveness

Both studies measured whether Alex Taylor would serve effectively in Congress. In study 1 predicted effectiveness declines from 3.27 at age 50 to 3.12 at age 70 and 2.97 at age 80. Relative to age 50, the age-70 and age-80 profiles are evaluated 0.149 and 0.306 points less favorably ($p < .001$ for both), whereas ages 30 and 50 do not differ ($p = .195$). The advanced-age penalty thus extends to perceived governing capacity.

Respondents retained meaningful, though imperfect, information about the treatment. Assigned and recalled age correlate at $r = .719$ in Study 1 and $r = .476$ in Study 2; 67% and 52%, respectively, recalled age within five years. On average, respondents recalled candidates as 1.95 years younger than assigned in Study 1 and 3.74 years younger in Study 2.

Age was not the dominant candidate cue. In the check-all-that-apply measure, 22.2% of Study 1 respondents and 25.8% of Study 2 respondents selected age as influential, compared with 61–63% selecting ideology and 43–44% selecting the candidate quotation. Its effect was instead concentrated at advanced ages, where evaluations and perceived effectiveness declined most clearly.

4.4 H4: Asymmetric age response

Figure 2 presents the pooled candidate-age response. Both studies show little change until middle age then followed by declining evaluations near the upper end of the assigned range. In Study 1, ages 30 and 50 do not differ significantly ($\Delta = 0.091$, $p = .069$). Relative to an otherwise comparable 50-year-old, a 70-year-old receives an evaluation 0.108 points lower and an 80-year-old an evaluation 0.236 points lower ($p < .001$ for both). Evaluations also decline from age 70 to 80 ($\Delta = -0.128$, $p < .001$). Study 2 reproduces the pattern on the 0–100 feeling thermometer. Ages 30 and 50 are statistically similar ($\Delta = 0.782$, $p = .516$), whereas 70- and 80-year-olds are evaluated 1.537 and 3.175 points less favorably than 50-year-olds ($p = .026$ for both). The decline from age 70 to 80 is also significant ($\Delta = -1.638$, $p = .041$).

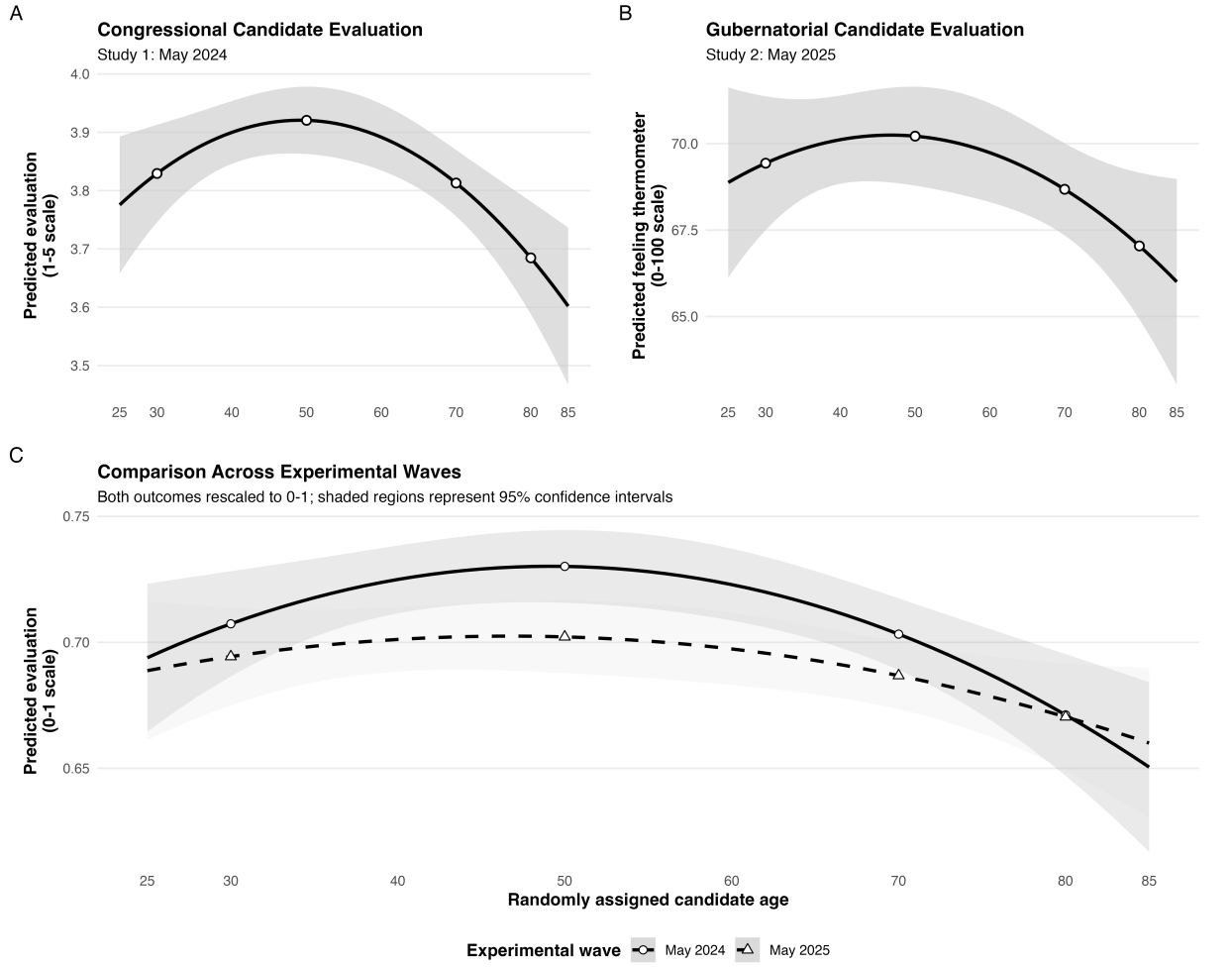

Figure 2: Predicted candidate evaluations across assigned age. Panels A and B report predictions in each study's native outcome units. Panel C places both outcomes on a common 0–1 scale. Shaded regions are 95% confidence intervals.

Table 3 provides several of these age contrasts. The age-80 versus age-50 contrast equals -0.257 standard deviations in Study 1 and -0.136 in Study 2. A descriptive inverse-variance synthesis yields -0.190 standard deviations (95% CI: $-0.280, -0.100$). Although the estimate is smaller in Study 2, the linear and quadratic age terms do not differ jointly across studies, $\chi^2(2) = 1.70$, $p = .427$.

Table 3: Selected candidate-age contrasts

Study	Contrast	Native difference	Difference in SDs	p
Spring 2024	Age 50 minus age 30	+0.091	+0.099	.069
Spring 2024	Age 70 minus age 50	−0.108	−0.117	$< .001$
Spring 2024	Age 80 minus age 50	−0.236	−0.257	$< .001$
Spring 2024	Age 80 minus age 70	−0.128	−0.140	$< .001$
Spring 2025	Age 50 minus age 30	+0.782	+0.034	.516
Spring 2025	Age 70 minus age 50	−1.537	−0.066	.026
Spring 2025	Age 80 minus age 50	−3.175	−0.136	.026
Spring 2025	Age 80 minus age 70	−1.638	−0.070	.041

Notes: Study 1 differences are measured on the 1–5 candidate-evaluation scale; Study 2 differences are feeling-thermometer points. Contrasts are average predicted differences from the quadratic specification with HC2 uncertainty.

These results support H1 and H2. Evaluations decline at ages 70 and 80 relative to age 50, while the difference between ages 30 and 50 is smaller and less reliable. The result is an advanced-age penalty averaged across the experience profiles, not a uniform linear decline or a comparable penalty at both ends of the age range.

Figure 3 compares quadratic and natural-spline predictions. Both show an upper-age decline, although their shapes differ at younger and middle ages. In Study 1, the spline performs better by AIC and the quadratic slightly better by BIC. In Study 2, the linear and quadratic specifications fit similarly, while the spline performs somewhat less well. The Study 2 quadratic term is not independently significant ($p = .115$), but the age-70 and age-80 contrasts remain significant. The robust finding is therefore the upper-age decline, not a single turning point or uniquely correct function.

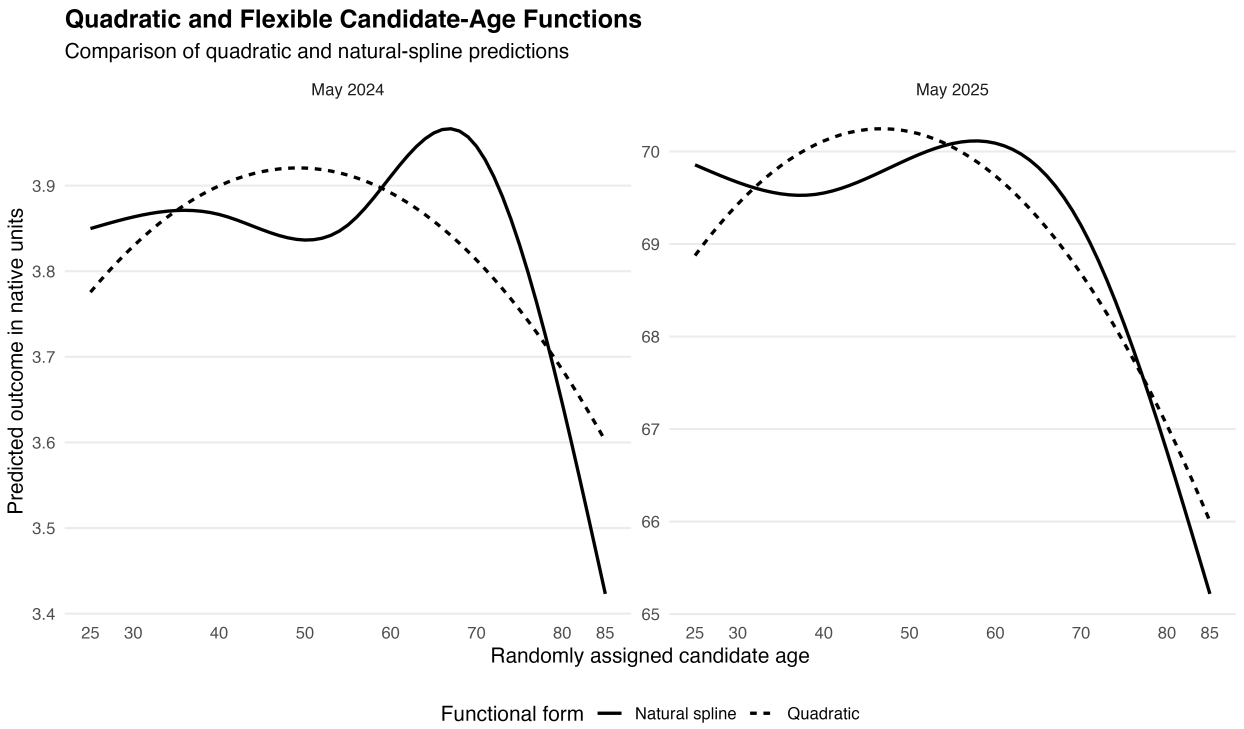

Figure 3: Quadratic and natural-spline candidate-age functions. Predictions are reported in each study's native outcome units.

Alternative outcome models lead to the same conclusion. An ordinal model for Study 1 produces a negative quadratic age term ($p = .004$). A fractional-logit model for the rescaled Study 2 outcome also indicates an upper-age decline, although its quadratic term is imprecise ($p = .117$). Finally, excluding insider profiles younger than age 30 leaves the Study 1 estimates essentially unchanged and strengthens the negative quadratic estimate in Study 2.[5] The advanced-age penalty is therefore not driven by profiles combining a very young age with multiple terms in office.

4.5 H5: Respondent-age moderation

Figure 4 presents predictions for respondents at the 20th, 50th, and 80th percentiles of age in each study. The Study 1 curves differ descriptively, but all decline at the oldest candidate ages. The linear and quadratic candidate-age interactions with respondent age are not jointly significant, $\chi^2(2) = 3.78$, $p = .151$. The Study 2 curves are more similar, and the joint interaction is effectively zero, $\chi^2(2) = 0.005$, $p = .998$. The advanced-age penalty is therefore neither confined to one voter generation nor explained by a preference for candidates near one's own age.

[5]See the potentially implausible young-insider analysis in Appendix D.

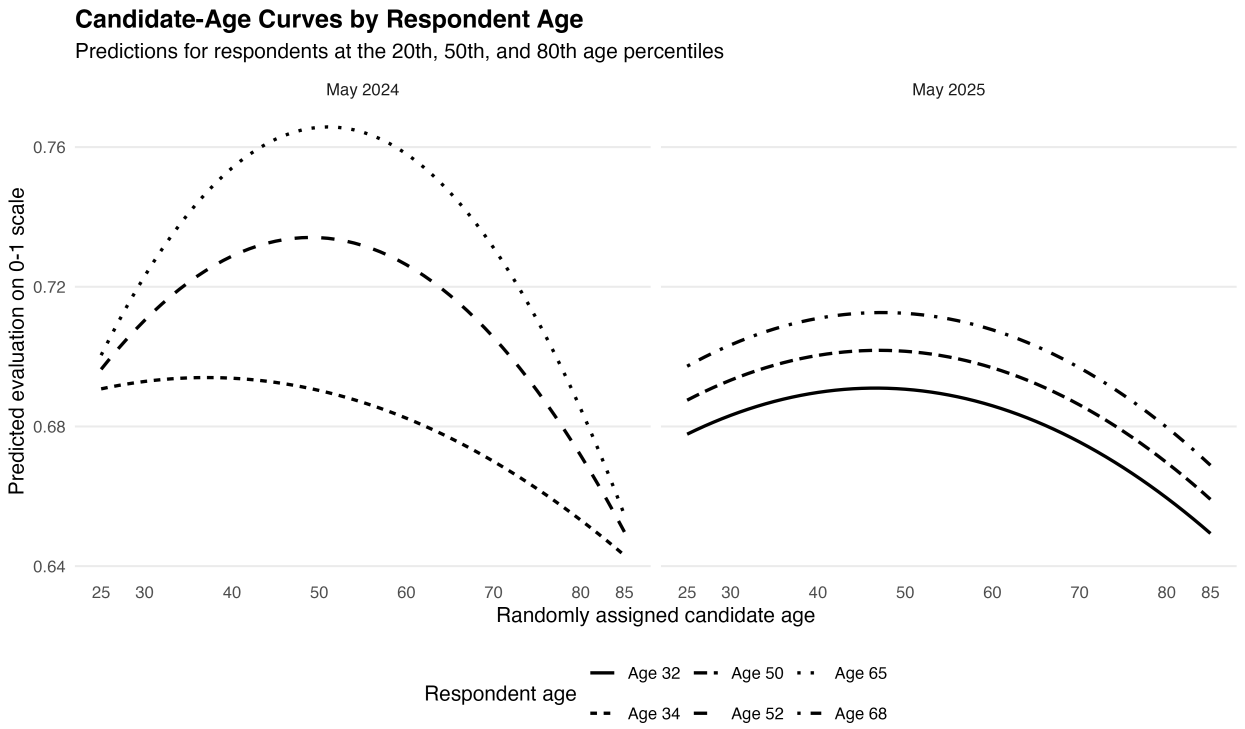

Figure 4: Candidate-age response by respondent age. Curves show predicted evaluations for respondents at the 20th, 50th, and 80th percentiles of the respondent-age distribution within each study.

5 How Candidate Age Changes Evaluative Language H6 & H7

The open-ended explanations provide secondary evidence about how respondents interpreted candidates of different ages. Because these responses were recorded after the closed-ended evaluation, they are analyzed as randomized post-treatment text outcomes rather than as measures of mediation. Models are estimated separately by study and average over the other randomized candidate attributes. Figure 5 shows changes in the language used to evaluate candidates as they age.

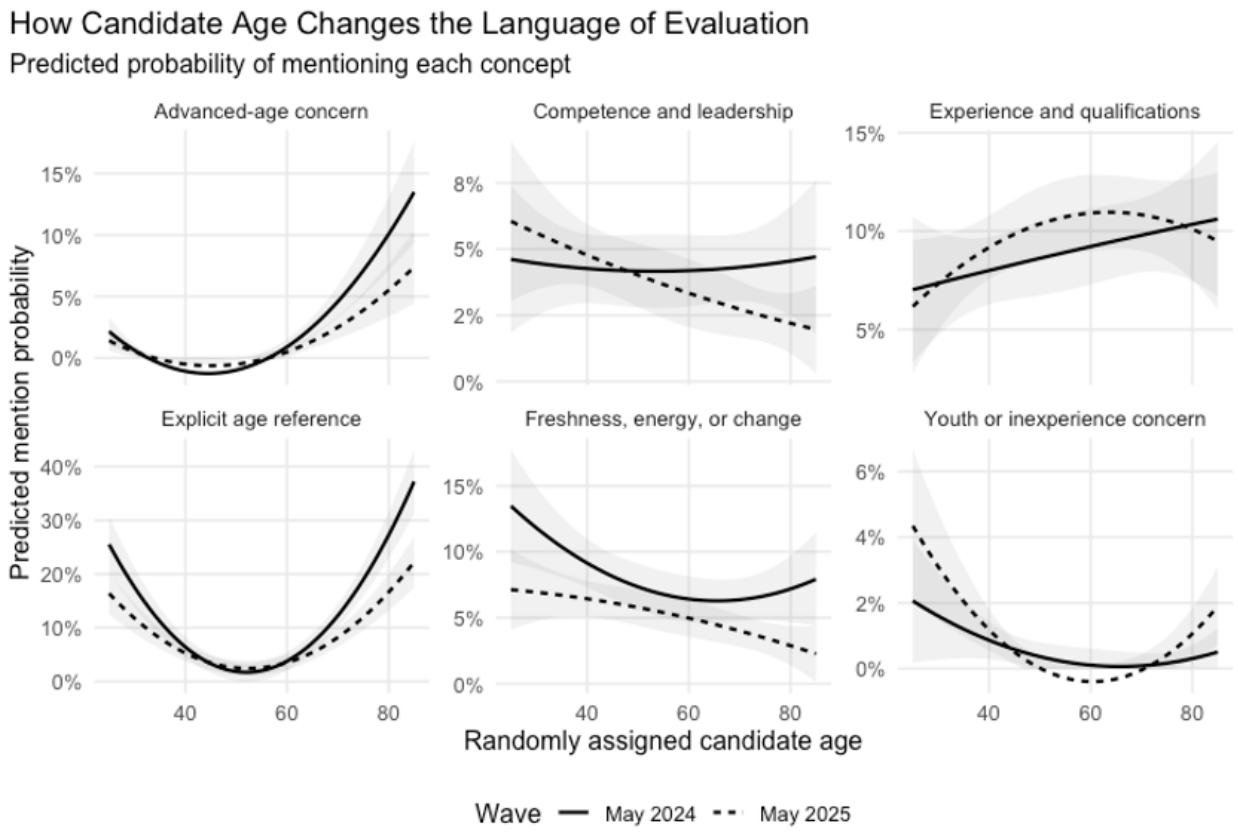

Figure 5: Candidate age and evaluative language. Curves report the predicted probability of an explicit age reference or an advanced-age concern by randomly assigned candidate age. Models are estimated separately by study and average over the other randomized candidate attributes. Shaded regions are 95% confidence intervals.

First, explicit references to candidate age are least common near the middle of the assigned range and rise sharply among the oldest profiles. In both studies, respondents evaluating candidates in their seventies and eighties were substantially more likely to mention age directly than respondents evaluating middle-aged candidates. This pattern supports H3: stated numerical age becomes especially salient at the upper end of the experimental range.

Second, advanced-age concerns become increasingly common as candidates approach ages 70 and 80. References to health, stamina, retirement, adaptability, capacity, or being out of touch are uncommon through most of the age range but increase among the oldest candidate profiles in both studies. This pattern supports H4 and corresponds to the same portion of the age range in which closed-ended evaluations decline.

The remaining theory-guided language categories do not display equally consistent patterns across studies and are reported Appendix E[6]. In particular, the results do not show a reliable

[6]See Appendix E for full evaluations of Open Ended Responses

compensating increase in references to experience or qualifications as candidate age rises. Supplementary lexical and out-of-wave supervised analyses provide convergent evidence: direct references to "age" and being "old" discovered in Study 1 reappear in Study 2, and text-implied evaluations become less favorable at advanced ages. As candidates enter advanced age, respondents become more likely to make age itself part of the evaluation and to raise concerns about the candidate's prospective capacity to serve.

6 Conclusion

The two experiments show that candidate age matters, but not in the simple form anticipated by every hypothesis. H1 received partial support: in Study 1, age produced the largest and only statistically detectable coefficient among the randomized demographic attributes, and the quadratic model revealed negative curvature. In Study 2, the age coefficients were negative but statistically imprecise, while some race effects were larger and statistically detectable. H2 was not supported as stated. Insider–outsider biographies shifted some baseline evaluations, but the direction varied by party and study, and the biographies did not detectably moderate the age-response function. H3 was also not supported. Respondents recalled candidate age with meaningful accuracy, but ideology and the candidate quotation were more often identified as influential.

The most consistent evidence supports H4. Evaluations changed little between ages 30 and 50 but declined at ages 70 and 80 in both studies. Relative to an otherwise comparable 50-year-old, an 80-year-old was evaluated 0.26 standard deviations less favorably in Study 1 and 0.14 standard deviations less favorably in Study 2. In Study 1, the same upper-age pattern appeared in perceived effectiveness, and the principal evaluation results remained under alternative specifications. The H5 moderation analysis found no detectable evidence that the age response varied with respondents' own ages. The exploratory open-ended expectations, H6 and H7, were supported. At the upper end of the assigned range, respondents became more likely to mention age directly and to express concerns about health, stamina, retirement, adaptability, capacity, or being out of touch. Because these responses followed the closed-ended ratings, they identify changes in expressed considerations rather than a causal mediation pathway. Future work could focus on the implications of this age penalty, especially in an era of increasing candidate age.

References

Cuddy, Amy J. C., Michael I. Norton, and Susan T. Fiske (2005). "This Old Stereotype: The Pervasiveness and Persistence of the Elderly Stereotype". In: *Journal of Social Issues* 61.2, pp. 267–285. DOI: 10.1111/j.1540-4560.2005.00405.x.

Fiske, Susan T. et al. (2002). "A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow from Perceived Status and Competition". In: *Journal of Personality and Social Psychology* 82.6, pp. 878–902. DOI: 10.1037/0022-3514.82.6.878.

Gramlich, John (Oct. 2023). *Most Americans Favor Maximum Age Limits for Federal Elected Officials, Supreme Court Justices*. Pew Research Center. URL: https://www.pewresearch.org/short-reads/2023/10/04/most-americans-favor-maximum-age-limits-for-federal-elected-officials-supreme-court-justices/ (visited on 07/16/2026).

Kirkland, Patricia A. and Alexander Coppock (2018). "Candidate Choice Without Party Labels: New Insights from Conjoint Survey Experiments". In: *Political Behavior* 40.3, pp. 571–591.

Lau, Richard R. and David P. Redlawsk (2001). "Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making". In: *American Journal of Political Science* 45.4, pp. 951–971. DOI: 10.2307/2669334.

Lawless, Jennifer L. and Richard L. Fox (2015). *Running from Office: Why Young Americans Are Turned Off to Politics*. New York: Oxford University Press.

Manning, Jennifer E. (Dec. 2024). *Membership of the 118th Congress: A Profile*. Tech. rep. R47470. Updated December 12, 2024. Congressional Research Service. URL: https://www.congress.gov/crs-product/R47470.

McClean, Charles T. and Yoshikuni Ono (2024). "Too Young to Run? Voter Evaluations of the Age of Candidates". In: *Political Behavior* 46, pp. 2333–2355. DOI: 10.1007/s11109-024-09920-2.

McDonald, Jared, David Karol, and Lilliana Mason (2020). ""An Inherited Money Dude from Queens County": How Unseen Candidate Characteristics Affect Voter Perceptions". In: *Political Behavior* 42.3, pp. 915–938.

North, Michael S. and Susan T. Fiske (2012). "An Inconvenienced Youth? Ageism and Its Potential Intergenerational Roots". In: *Psychological Bulletin* 138.5, pp. 982–997. DOI: 10.1037/a0027843.

Popkin, Samuel L. (1991). *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. Chicago: University of Chicago Press.

Riggle, Ellen D. B. and Marcia M. S. Johnson (1996). "Age Differences in Political Decision-Making: Strategies for Evaluating Political Candidates". In: *Political Behavior* 18.1, pp. 99–118.

Roberts, Damon C. and Jennifer Wolak (2023). "Do Voters Care About the Age of Their Elected Representatives?" In: *Political Behavior* 45.4, pp. 1959–1978.

Shames, Shauna L. (2017). *Out of the Running: Why Millennials Reject Political Careers and Why It Matters*. New York: New York University Press.

Sigelman, Lee and Carol K. Sigelman (1982). "Sexism, Racism, and Ageism in Voting Behavior: An Experimental Analysis". In: *Social Psychology Quarterly* 45.4, pp. 263–269.

Stockemer, Daniel and Aksel Sundström (2023). "Age Inequalities in Political Representation: A Review Article". In: *Government and Opposition*.

Woods, Neal D., Barry L. Tadlock, and Ellen D. B. Riggle (2003). "Getting Older, Getting Wiser? The Impact of Age on Candidate Evaluation". In: *Politics & Policy* 31.3, pp. 432–451.