hexagon:2611.00825v1

Reconstructing the Simulmatics 1960 Simulation: Mathematics, Methods and a Partial Replication

John Hessler
Applied Mathematician & Computer Scientist
Relic {Code} Lab
jhessle1@jhu.edu

21 November 2026

Abstract

In August 1960 the Simulmatics Corporation ran a computer simulation for the Democratic National Committee. It ranked 32 non-Southern states by Kennedy's likely strength if religion dominated the campaign. Pool and Abelson (1961) reported that this ranking correlated .82 with the November result, using no political data later than October 1958. This paper reconstructs the mathematics behind that result from the published record: the 1961 *Public Opinion Quarterly* article and the list of 480 voter types later printed in the appendix of Burdick's *The 480* (1964).

We make four contributions. First, we write the model as a complete system: a 480-type typology, a pooled issue-cluster estimator, a synthetic-state aggregator and nine cross-pressure equation sets. Second, we show that the model's one non-trivial step, estimating how many 1958 Democrats were anti-Catholic, is an odds-ratio constrained 2×2 table with a unique feasible solution for every odds ratio, and we derive its comparative statics. Third, we prove that the authors' "lucky guess" benchmark (add 34 per cent of the Catholic share to the 1956 Democratic vote) is the in-sample ratio of two least-squares coefficients, so its correlation is a two-parameter multiple correlation fitted with hindsight. Fourth, we recompute the published benchmarks from public returns. The authors' printed ranking has a Spearman correlation of 0.755 with Kennedy's 1960 share across the 32 states ($p < 10^{-6}$). The model's own list of failures, the Mountain states, reappears in our data. Six animations, collected in an online story map, accompany the paper.

On the contrary, we can be sure that they are partly false.

— Ithiel de Sola Pool and Robert Abelson, "The Simulmatics Project" (1961, p. 176)

1. Introduction

The Simulmatics project is usually remembered as a story about politics and computing: a small company, time on the MIT Computation Center's computer, and a candidate who won by a narrow margin. It is less often read as mathematics. Yet the 1961 report by Ithiel de Sola Pool and Robert Abelson contains an explicit, testable model of voting behaviour. It specifies inputs, equations, parameters and an aggregation rule, and it reports how well the output matched the election.

The model's purpose was narrow. It did not forecast the 1960 result. It answered a contingent question: if the campaign were fought on party and Catholicism alone, which states would Kennedy do best in, and which worst? The answer was a ranking of 32 states, delivered to the campaign on

25 August 1960 together with two other reports (Pool and Abelson 1961, p. 173).

This paper asks three questions of that model.

1. What exactly did it compute? We rebuild the full system from the article's equations and the published list of voter types.
2. Why did it work as well as it did? We analyse each component, the fourfold-table step in particular, and the authors' argument that working voter type by voter type adds information a single correction cannot.
3. How much of the reported performance can be checked today? We recompute every benchmark that public data allow and set out what is needed for the rest.

The contributions are a formal statement of the model (Sections 3–6), three analytical results (Sections 7–9), a partial replication (Section 10) and a reusable code base with an animated visualisation (Section 11). Section 12 sets out the limits of what can be recovered and a research agenda.

2. Sources and data

Three historical sources define the model; three modern ones are used to test it.

The 1961 article. Pool and Abelson describe the data, the 480-by-52 matrix, the synthetic states and the simulation, and print the equations for all nine cross-pressure groups, seven of them in footnote 8 (pp. 177–179). They also print the 32-state ranking (pp. 176–177) and a table of benchmark correlations (p. 182).

The 480 appendix. The full list of voter types was printed in the appendix of Eugene Burdick's novel *The 480* (1964, pp. 311–313), with the company's permission. It gives the defining attributes of every type, the community-size thresholds and the assignment of states to regions.

The underlying archive. Simulmatics pooled 66 surveys held by the Roper Public Opinion Research Center: 50 from the 1952–1958 elections (about 85,000 respondents) and 16 anticipating 1960, well over 100,000 interviews in all. The tabulated matrix was deposited with the Roper Center as a condition of access; the simulation programs and census-based state weights were not (p. 171). Neither the matrix nor the weights is printed in the article.

Data used in this paper.

Series	Source	Use
1956 and 1960 presidential returns by state	Official state totals, via Wikipedia's election pages	Benchmarks; outcome variable
Simulmatics ranking of 32 states	Pool and Abelson (1961, pp. 176–177)	Tested against 1960 returns
Voter typology and state regions	Burdick (1964, pp. 311–313)	Model structure; state aggregation

Series	Source	Use
Catholic share of state population	Gallup self-identification, pooled 2000–2004	Placeholder only (Section 10.3)

The returns were checked for internal consistency: every published percentage matches votes divided by total to within 0.005 points, except Alabama 1960, where the unpledged-elector slate breaks the usual arithmetic. Alabama is outside the 32-state sample. The Catholic-share series is a stated placeholder. Its proper replacement, the National Council of Churches' 1952 state enumeration, is discussed in Section 12.

3. The voter typology

The electorate is partitioned into 480 voter types. A type is a cell in a cross-classification of party, region, religion or ethnicity, sex, community size and occupational status. The partition is not a full Cartesian product: cells were merged where samples were thin, which is why the total is 480 rather than a round power of the factor levels.

3.1 Structure

The four main regions share one 36-cell template, repeated for each of three party identifications. The Border states use a smaller 16-cell template without community size.

$$|\mathcal{T}| \;=\; \underbrace{4 \times 3 \times 36}_{\text{East, South, Midwest, West}} \;+\; \underbrace{1 \times 3 \times 16}_{\text{Border}} \;=\; 432 + 48 \;=\; 480$$

Block	Cells	Defining attributes
Protestant, by sex	$2 \times 7 = 14$	Community × occupation (below)
Catholic, by sex	$2 \times 7 = 14$	Community × occupation (below)
Jewish	2	Professional and white collar; blue collar
Black	2	Professional and white collar; blue collar
Protestant, no SES rating	1	Non-labour-force families
Catholic, no SES rating	1	Non-labour-force families
Other religion	1	—
No religion	1	—
Total per region × party	**36**	

Block	Cells	Defining attributes

The seven community × occupation cells are: rural and town, each split into professional-and-white-collar and blue collar; and urban (over 100,000), split three ways into professional, white collar and blue collar. Rural means under 5,000 inhabitants and town 5,000 to 100,000. Professional and white-collar workers were not separated outside cities because rural areas and towns had too few professionals. Types are numbered region by region: 1–108 East, 109–216 South, 217–324 Midwest, 325–432 West and 433–480 Border. Within each region the parties run Democrat, Republican, Independent.

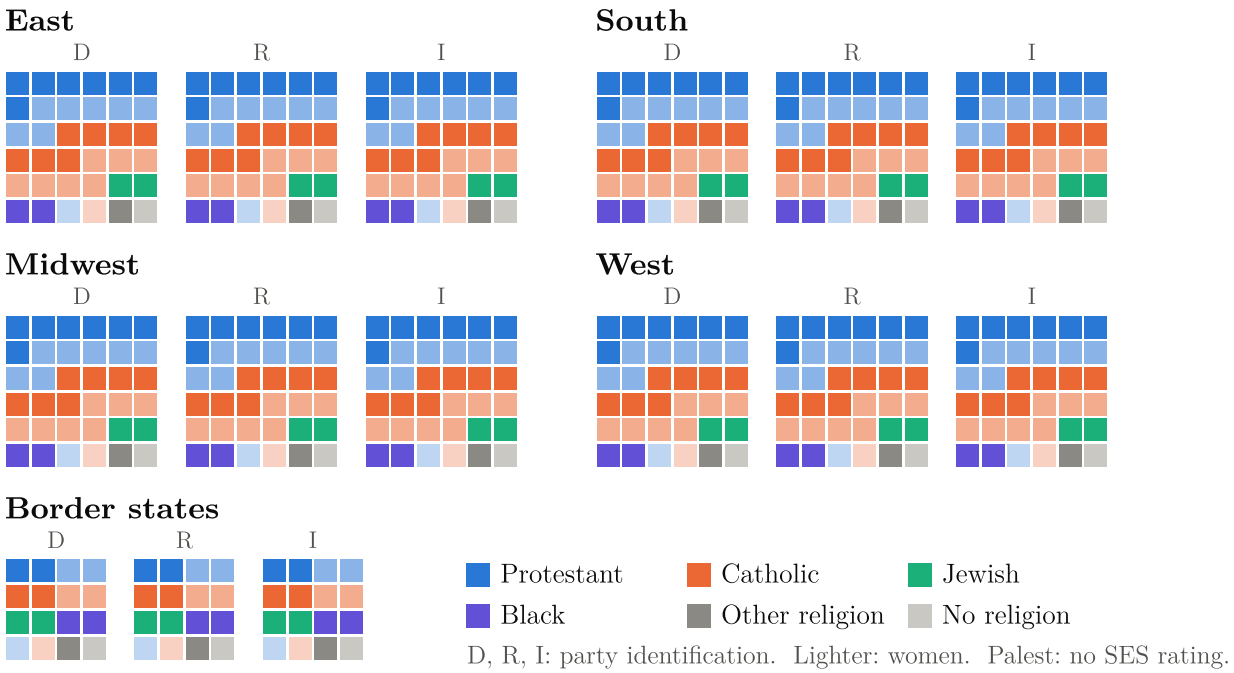


Figure 1: The 480 voter types. One 36-cell template (Protestants and Catholics by sex, community size and occupation; two Jewish and two Black cells; no-SES, other-religion and no-religion cells) is repeated for three party identifications (D, R, I) in each of four regions, giving 432 types. The Border states add 3 × 16 types without a community-size split. A state draws only on its own region's 108 types. Source: Burdick (1964, pp. 311–313). Animated version: Animation S1 in the online story map.

3.2 Regions

The East has 9 states (Maine to Pennsylvania), the South 11 (Virginia to Texas, including Oklahoma), the Midwest 12 (Ohio to Kansas, including Missouri), the West 11 (Montana to Washington) and the Border 5 (Maryland, Delaware, West Virginia, Kentucky, Tennessee). There is no Mountain region. Arizona, Colorado, Idaho, Nevada, New Mexico, Utah and Wyoming are pooled with the Pacific coast.

3.3 Consistency with the 1961 article

Three independent checks connect the appendix to the article. First, the article states that a state contains at most 108 voter types (p. 175); the template gives exactly $3 \times 36 = 108$ per main region. Second, footnote 5 blames the worst misses on "the absence of a 'Mountain Region' in our classification", which the region list confirms. Third, the 32 states the article ranks are exactly the $9 + 12 + 11$ states of the East, Midwest and West: South and Border are excluded, and no state is missing or extra. We verified this set identity in code.

3.4 Coding rules and anomalies

The appendix adds two rules that matter for estimation. Types marked with one star do not include urban white-collar respondents, while types marked with two stars do. The stars fall on the Jewish, Black and Border-state types: as printed, urban white-collar respondents in those groups are counted with blue collar, not with professional and white collar. We record this as printed.

One entry is irregular. Type 441 (Border, Democrat, Jewish, professional and white collar) is printed with "Female" in the sex column, although no other Jewish type has a sex. We treat it as a typesetting carry-over from type 439.

Finally, the Black types carry no religion, but the simulation's cross-pressure equations need one (Section 6). Most Black voters in 1960 were Protestant, so our default maps these types to the Protestant equations. Because this choice is ours, not the authors', the code exposes it as a parameter.

4. Issue clusters and the data matrix

The survey archive was reduced to a matrix with one row per voter type and one column per "issue cluster": 480 rows by 52 columns. Each cell holds four numbers: the count of respondents asked, and the shares pro, anti and undecided. A separate matrix was built for each biennial election (1952, 1954, 1956, 1958) and one consolidated matrix for all years, so trends could be read by comparing matrices.

4.1 Trichotomisation

Original codes, sometimes up to thirteen per question, were collapsed to three outcomes. For a question with an ordered scale, the authors set cutting points "depending on the distribution" (p. 169). Formally, for question q with ordered codes $1, \ldots, m$, a pair of cut points $c_1 \leq c_2$ assigns code y to anti if $y \leq c_1$, to pro if $y > c_2$ and to undecided otherwise. The choice of cut points is a researcher degree of freedom that the article does not report.

4.2 The pooled cell estimator

Let Q_j be the set of questions assigned to cluster j, and let x^{+}_{tq} and n_{tq} be the pro replies and cards for type t on question q. The cell value pools replies across every question and survey in the cluster:

$$\hat{p}^{+}_{tj} \;=\; \frac{\sum_{q\in Q_j} x^{+}_{tq}}{\sum_{q\in Q_j} n_{tq}} \;=\; \sum_{q\in Q_j} \frac{n_{tq}}{n_{tj}}\, p^{+}_{tq}, \qquad n_{tj} = \sum_{q\in Q_j} n_{tq}$$

The estimator is a sample-size weighted average of question-level proportions. It is unbiased for a common underlying proportion only if the questions in the cluster measure the same attitude with the same marginal level. Different wordings with different levels add a component of bias that depends on which questions happened to reach which types.

The pooling is also what makes small groups visible. With about 100,000 interviews spread over 480 types, a type averages roughly 210 respondents. A cluster asked in only a tenth of the surveys leaves about 21 per cell, and a sampling standard error near $\sqrt{0.25/21} \approx 0.11$ at $p = 0.5$. This is the reason the authors built synthetic states rather than reading each state's own respondents (Section 5).

4.3 Correlational validation of clusters

Whether two questions belong in one cluster was decided in two steps: an a priori grouping, then an empirical test. For the test the 480 types were consolidated into 15 larger groups, $g = 1, \ldots, 15$, so that single-survey percentages were stable. For each pair of candidate questions the authors correlated the pro shares across the 15 groups, and separately the anti shares:

$$\rho^{+}_{qq'} = \mathrm{corr}_g\left(p^{+}_{gq},\, p^{+}_{gq'}\right), \qquad \rho^{-}_{qq'} = \mathrm{corr}_g\left(p^{-}_{gq},\, p^{-}_{gq'}\right)$$

Questions were retained only when both correlations were high and the items also seemed politically equivalent. No threshold is reported.

The second condition was essential, and a simple model shows why. Suppose each question's pro share in group g is a linear function of the group's Democratic share π_g plus noise: $p^{+}_{gq} = \alpha_q + \beta_q \pi_g + \varepsilon_{gq}$. Then any two strongly party-linked questions correlate highly across groups, whatever their content. The authors give exactly this example: "Which party is better for people like you?" and a question on Adlai Stevenson's image correlate strongly but do not measure the same attitude (p. 170). Correlation across 15 groups is evidence that two questions sort groups the same way. It is not evidence that they measure one attitude.

The validation failed for several foreign-policy clusters, where answers were "highly sensitive to conjunctural influences or shades in wording" (footnote 2). Those were kept as unvalidated a priori clusters and, by the authors' account, barely used. The test also detected a change over time: replies to an identical question on Senator McCarthy stopped tracking the McCarthyism cluster after his censure.

5. Synthetic states

A state's predicted behaviour is the population-weighted average of the behaviour of the voter types living in it. With about 2,000 interviews per state on average, and as few as 300 to 400 in small

states, direct state estimates were too noisy. The authors instead estimated how many persons of each voter type lived in each state, from census, poll and voting data, and borrowed the type's behaviour from every respondent of that type in the region.

Let T(s) be the voter types of state s (at most 108, all from s's region), w_{st} the estimated number of persons of type t in s, and V_t any type-level quantity. The synthetic state is

$$\bar{V}_s \;=\; \frac{\sum_{t \in T(s)} w_{st}\, V_t}{\sum_{t \in T(s)} w_{st}}$$

The estimator rests on a homogeneity assumption, which the authors state plainly and call "partly false" (p. 176): a voter type behaves identically in every state of its region. An upper-income rural Protestant Republican man is the same voter in Maine and in New York. All differences between states within a region therefore come from their mix of types, not from place.

In modern terms this is post-stratification without the multilevel model: cell estimates pooled within region, reweighted to each state's composition. It is the ancestor of multilevel regression and post-stratification (MRP). The difference is that MRP shrinks each cell towards a fitted model, partly pooling across cells, while Simulmatics pooled completely across states within a region and not at all across types.

The assumption is testable, which is what made the project scientifically interesting to its authors. If the 1960 result is well predicted, the variation between states within a region is mostly compositional. Where it fails, place matters. The authors found the largest failures in Arizona, Nevada, New Mexico, Idaho and Colorado, mostly small Mountain states that the classification forced into a West dominated by California (footnote 5). Section 10 shows that the same states stand out in our recomputation.

The weights w_{st} are the binding constraint on any replication. The census can supply each state's population by sex, community size, occupation and race, but religion and party are not census items. They must come from survey distributions within each region, conditional on the census attributes. Simulmatics did this before any 1960 census results were available, presumably from 1950 figures, and its weights were never published.

6. The cross-pressure equations

The simulation imagines a campaign in which the only issues are party and Catholicism. It draws on the cross-pressure finding of Berelson, Lazarsfeld and McPhee (1954): voters whose group loyalties pull in opposite directions defect more and vote less. The 480 types are sorted into a 3×3 grid of religion (Protestant, Catholic, other) by party (Republican, Democrat, Independent). Each cell gets its own pair of equations, judged by the analysts from what the data showed about that kind of voter.

6.1 Notation

All quantities are proportions for one voter type. The subscripts mix two conventions, which is a trap for readers: 56 and 58 are election years, while 35 and 14 are issue-cluster numbers.

Symbol	Meaning
P_{56}, Q_{56}	Share preferring Stevenson / Eisenhower in 1956 polls
P_{58}, Q_{58}	Share intending to vote Democratic / Republican in 1958 polls
P_{35}	Nonvoting index (issue cluster 35)
P_{14}	Share giving anti-Catholic replies (issue cluster 14)
V_k, V_n	Predicted Kennedy / Nixon share of the type in 1960
a	Anti-Catholic 1958 Democrats, from the fourfold table (Section 7)

Because replies were trichotomised, $P + Q < 1$: the undecided remainder is never reallocated. V_k and V_n therefore do not sum to the type's turnout, and the state figure is what the authors call a "Kennedy index", not strictly a percentage.

6.2 The nine cells

Protestant Republicans: no cross-pressure. Their 1960 vote is expected to repeat 1956, discounted by their record of nonvoting:

$$V_k = P_{56}\,(1 - P_{35}), \qquad V_n = Q_{56}\,(1 - P_{35})$$

Protestant Democrats and Protestant Independents: cross-pressured. The base is 1958, because too many of them defected to Eisenhower in 1956 for that year to show normal behaviour. The anti-Catholic Democrats, a, move to Nixon; and the nonvoting index is doubled, because cross-pressured voters stay home more:

$$V_k = (P_{58} - a)\,(1 - 2P_{35}), \qquad V_n = (Q_{58} + a)\,(1 - 2P_{35})$$

Catholic Democrats and Catholic Independents. A third of those who voted Republican in 1958 return to the Democrats; turnout is not penalised:

$$V_k = \Big(P_{58} + \tfrac{1}{3}Q_{58}\Big)(1 - P_{35}), \qquad V_n = \tfrac{2}{3}\,Q_{58}\,(1 - P_{35})$$

Catholic Republicans: cross-pressured the other way. The same one-third switch, with a doubled nonvoting penalty on the Kennedy side and a tripled one on the Nixon side:

$$V_k = \Big(P_{58} + \tfrac{1}{3}Q_{58}\Big)(1 - 2P_{35}), \qquad V_n = \tfrac{2}{3}\,Q_{58}\,(1 - 3P_{35})$$

All others (Jewish, other religion, no religion; all parties). The base averages 1956 and 1958, and the fourfold step is applied to that average:

$$V_k = \Big(\tfrac{P_{58}+P_{56}}{2} - a\Big)(1 - 2P_{35}), \qquad V_n = \Big(\tfrac{Q_{58}+Q_{56}}{2} + a\Big)(1 - 2P_{35})$$

With the margins of the fourfold table set to a + b = $(P_{58} + P_{56})/2$, c + d = $(Q_{58} + Q_{56})/2$ and a + c = $P_{14}\times$ $(P_{58} + Q_{58} + P_{56} + Q_{56})/2$.

6.3 Properties of the system

Three features of the equations are worth stating explicitly.

Conservation within the decided. Every equation moves voters between the two candidates without creating or destroying them. Before turnout, $V_k + V_n$ equals $P_{58} + Q_{58}$ in the Protestant and Catholic Democratic cells, so the anti-Catholic step and the one-third return are pure transfers. The Catholic Republican cell breaks this symmetry only through its unequal turnout factors.

Domain constraints. The turnout factors are linear in P_{35}, so they become negative for heavy nonvoters: $1-2P_{35} < 0$ when $P_{35} > 0.5$, and $1-3P_{35} < 0$ when $P_{35} > 1/3$. A type with a nonvoting index above one third would receive a negative Nixon vote in the Catholic Republican cell. Any implementation must check these bounds; ours clips nothing and tests that outputs stay in [0, 1] for realistic inputs.

Judgemental parameters. Four constants enter every calculation: the odds ratio 0.6, the return fraction 1/3, and the turnout multipliers 2 and 3. The odds ratio was averaged from the few surveys that asked both questions. The one-third was chosen after reading Kennedy–Nixon trial heats, and the multipliers rest on the cross-pressure literature rather than on an estimate. The authors acknowledge that they did not explore which parameters drove the result (p. 180); Section 12 proposes that exercise.

7. The fourfold-table problem

The model's only non-trivial computation answers one question: of the Democrats who gave anti-Catholic replies, how many had already voted Republican in 1958? Those "hopeless cases" cost Kennedy nothing new. Only anti-Catholics still voting Democratic in 1958 are net losses.

The obstacle was that no respondent answered both questions in most surveys. One respondent of a type was asked his 1958 vote; another, on a different survey, was asked whether he would vote for a Catholic. The joint distribution had to be reconstructed from its margins plus one measure of association.

7.1 Formulation

Within a voter type, cross 1958 vote intention with anti-Catholicism:

1958 vote intention	Anti-Catholic	Not anti-Catholic	Row total
Democratic	a	b	$R = P_{58}$
Republican	c	d	Q_{58}
Column total	$C = P_{14}N$	$N - C$	$N = P_{58} + Q_{58}$

The margins are known for each type. The association was measured on the few surveys carrying both questions: across the larger Protestant Democrat types, the cross-product ratio averaged about 0.6. The unknown a therefore satisfies

$$a + b = R, \qquad a + c = C, \qquad a + b + c + d = N, \qquad \frac{ad}{bc} = \theta = 0.6$$

Substituting $b = R - a$, $c = C - a$ and $d = N - R - C + a$ into the odds-ratio condition gives a quadratic in a:

$$(1 - \theta)\, a^2 \;+\; \left[\, N - R - C + \theta\,(R + C) \,\right] a \;-\; \theta\, R\, C \;=\; 0$$

The article says only that a "was arrived at by the computer". The quadratic is what that computation must have solved.

7.2 Existence and uniqueness

Proposition 1. For any margins with $0 < R, C < N$ and any odds ratio $\theta > 0$, the system has exactly one solution with all four cells non-negative.

Proof. All cells are non-negative exactly when a lies in the interval $[L, U]$, with $L = \max(0, R+C-N)$ and $U = \min(R, C)$. Define $g(a) = a\, d(a) - \theta\, b(a)\, c(a)$. On $[L, U]$, a and $d(a)$ are non-negative and increasing in a, while $b(a)$ and $c(a)$ are non-negative and decreasing. So g is strictly increasing. At $a = L$, either $a = 0$ or $d = 0$, so $g(L) = -\theta bc \le 0$. At $a = U$, either $b = 0$ or $c = 0$, so $g(U) = ad \ge 0$. By continuity g has exactly one root in $[L, U]$. □

In practice: when $\theta < 1$ the quadratic's leading coefficient is positive and its constant term negative, so its roots have opposite signs and the feasible root is the positive one. When $\theta = 1$ the equation is linear.

7.3 The independence benchmark

At $\theta = 1$ the solution is $a = RC/N = P_{14}P_{58}$. The share of Democrats lost equals the share of the type that is anti-Catholic. This is exactly the naive correction the authors rejected: "it would not suffice merely to subtract the percentage who gave anti-Catholic replies" (p. 178). The odds ratio is

therefore the single parameter that measures how far the model departs from naive subtraction. A value below 1 says anti-Catholics were already disproportionately Republican, so fewer Democratic votes are at risk.

7.4 Comparative statics

Implicit differentiation of g(a; θ, C) = 0 gives the two derivatives that drive the model's sensitivity:

$$\frac{\partial a}{\partial \theta} = \frac{b\,c}{a + d + \theta\,(b + c)} > 0, \qquad \frac{\partial a}{\partial C} = \frac{a + \theta\,b}{a + d + \theta\,(b + c)} \in (0, 1)$$

Since $V_k = (R - a)(1 - 2P_{35})$ and $C = P_{14}N$, a Protestant Democrat type's Kennedy share falls with anti-Catholicism at the rate

$$\frac{\partial V_k}{\partial P_{14}} = -\,(1 - 2P_{35})\,N\,\frac{a + \theta\,b}{a + d + \theta\,(b + c)}$$

The loss per point of anti-Catholicism is less than one point of the decided vote. It shrinks as θ falls, because more of the anti-Catholics are found among existing Republicans.

7.5 Worked example

Take an illustrative Protestant Democrat type with $P_{58} = 0.55$, $Q_{58} = 0.35$, $P_{14} = 0.30$ and $P_{35} = 0.20$. These are constructed values chosen for clarity, not archive data.

θ	a	b	c	d	V_k	V_n
0.3	0.1105	0.4395	0.1595	0.1905	0.2637	0.2763
0.6 (as used)	0.1417	0.4083	0.1283	0.2217	0.2450	0.2950
1.0 (naive subtraction)	0.1650	0.3850	0.1050	0.2450	0.2310	0.3090
1.5	0.1828	0.3672	0.0872	0.2628	0.2203	0.3197

Every row reproduces its odds ratio to machine precision. Moving from naive subtraction to the authors' $\theta = 0.6$ raises this type's Kennedy index by 1.4 points, about 6 per cent of its value. At $\theta = 0.6$ the derivative $\partial a/\partial\theta$ is 0.076: a 0.1 error in the odds ratio moves the type's Kennedy index by about 0.5 points after the turnout factor.

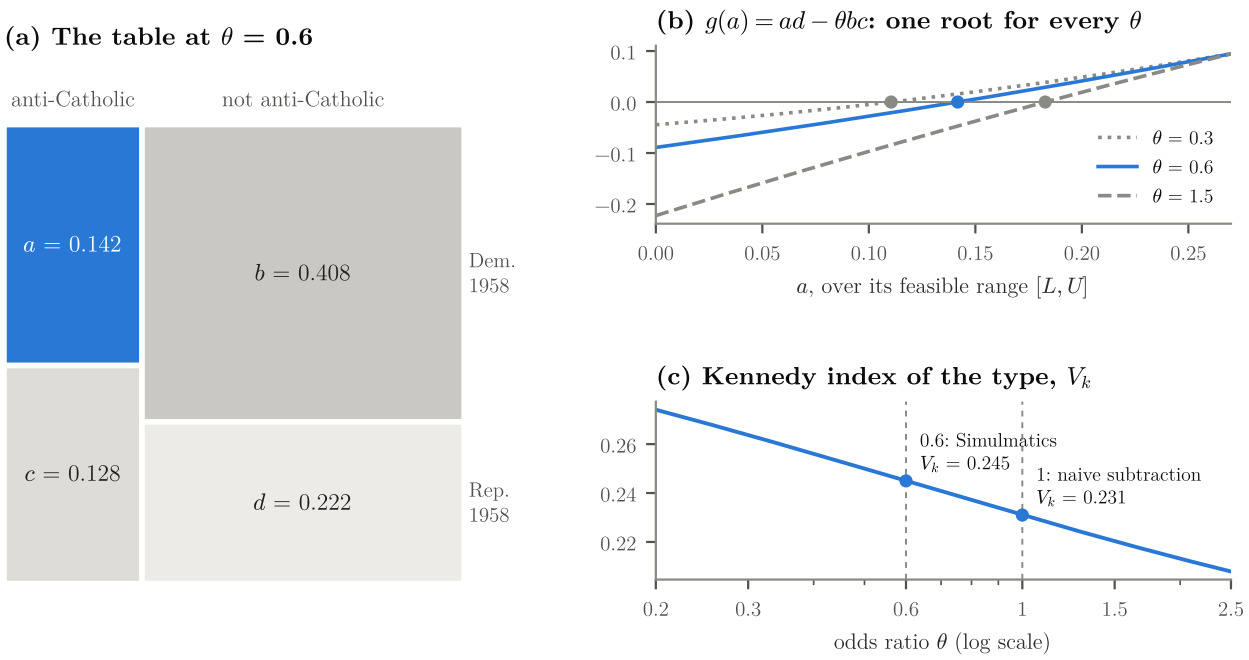


Figure 2: The fourfold-table step for an illustrative Protestant Democrat type ($P_{58} = 0.55$, $Q_{58} = 0.35$, $P_{14} = 0.30$, $P_{35} = 0.20$). (a) With the margins fixed, the authors' odds ratio $\theta = 0.6$ gives $a = 0.142$ anti-Catholic 1958 Democrats. (b) $g(a)$ rises across the feasible range, so every θ has exactly one root (Proposition 1). (c) The type's Kennedy index falls as θ rises; $\theta = 1$ reproduces naive subtraction. Animated version: Animation S2 in the online story map.

8. Why voter-type modelling differs from a global correction

Computing type by type changes the answer only when the model is non-linear and its inputs vary together across types. Pool and Abelson make this argument informally (p. 182): if the same process operated in every type, adding 34 per cent of each Catholic type to the Democrats would equal adding 34 per cent of all Catholics to the total. Interactions, such as turnout varying alongside party loyalty, break the equality. This section makes the argument exact.

8.1 Statement

Let x_t be the input vector of type t (P_{56}, Q_{56}, P_{58}, Q_{58}, P_{35}, P_{14}), f the equation for its cell, and w_t its weight in a state, with weights summing to one. The simulation computes the left-hand side below; a global correction computes something like the right-hand side, the model applied to the state's average voter.

$$\Delta_s \;=\; \sum_t w_t\, f(x_t) \;-\; f\Big(\sum_t w_t\, x_t\Big)$$

Proposition 2. $\Delta_s = 0$ for every set of weights and inputs if and only if f is affine.

This is Jensen's equality case. Every cell equation fails it, because each multiplies a vote share by a turnout factor.

8.2 The bilinear terms are covariances

For the bilinear cells the gap is exactly a weighted covariance. For Protestant Republicans:

$$\sum_t w_t \, P_{56,t} \, (1 - P_{35,t}) \;=\; \bar{P}_{56} \, (1 - \bar{P}_{35}) \;-\; \mathrm{Cov}_w(P_{56},\, P_{35})$$

So working type by type subtracts the covariance between Stevenson support and nonvoting. In the cross-pressured cells the turnout factor is $1 - 2P_{35}$ and the covariance term doubles. The fourfold step adds genuine curvature, since a is a square-root function of the margins. To second order its contribution is $\frac{1}{2}\,\mathrm{tr}(H\Sigma)$, where H is the Hessian of the cell equation and Σ the weighted covariance matrix of the inputs.

8.3 An illustration

Take two illustrative Protestant Democrat types of equal weight. Type A is loyal and participatory: $P_{58} = 0.70$, $Q_{58} = 0.20$, $P_{35} = 0.10$, $P_{14} = 0.10$. Type B is marginal: $P_{58} = 0.40$, $Q_{58} = 0.50$, $P_{35} = 0.35$, $P_{14} = 0.45$. Computed separately and averaged, their Kennedy index is 0.292. Applied once to their averaged inputs, the same equation gives 0.231. Of the 6.1-point gap, 4.9 points are the covariance term: the most Democratic type also votes most. The remaining 1.2 points come from the curvature of the fourfold step.

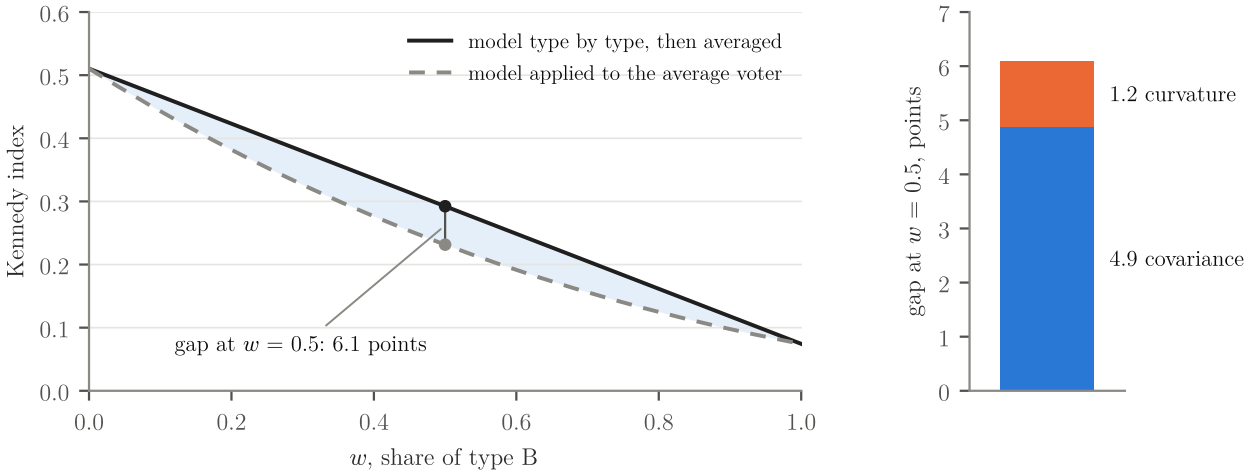


Figure 3: Type-by-type versus average-voter aggregation for the two illustrative Protestant Democrat types of Section 8.3, mixed in shares $1 - w$ and w. The two methods agree only at $w = 0$ and $w = 1$. At $w = 0.5$ the gap is 6.1 points: 4.9 from the covariance of vote and turnout across types and 1.2 from the curvature of the fourfold step. Animated version: Animation S3 in the online story map.

8.4 What this implies for the 1960 result

The type-by-type machinery can only beat a well-chosen global correction where these covariances differ from state to state. Composition effects that are the same everywhere shift every state's index equally and leave the ranking untouched. The authors' own comparison is consistent with that: their simulation (.82) and the best hindsight correction (.83) correlate with the outcome almost

equally. Their claim is therefore not that the simulation was more accurate than the best simple rule. It is that the simulation reached that accuracy before the election, from explicit assumptions, while the simple rule needed a parameter that only hindsight could supply. Section 9 makes the second half of that claim precise.

9. The lucky-guess benchmark

The authors' most quoted comparison is against a simple rule: add k times the state's Catholic percentage to its 1956 Democratic vote. With k = 0.34 the rule correlates .83 with the 1960 result, slightly better than the simulation's .82. No other k does better, and at k = 0.32 or 0.40 the correlation falls below .80 (p. 182). This section shows what that optimum is.

9.1 The correlation as a function of k

Write S for the 1956 Stevenson percentage, C for the Catholic percentage and K for the 1960 Kennedy percentage, all across the 32 states, with variances σ^2 and covariances σ subscripted in the obvious way. Then

$$r(k) \;=\; \operatorname{corr}(S + kC,\; K) \;=\; \frac{\sigma_{SK} + k\,\sigma_{CK}}{\sigma_K \sqrt{\sigma_S^2 + 2k\,\sigma_{SC} + k^2\sigma_C^2}}$$

Setting $r'(k) = 0$ gives a single interior stationary point:

$$k^* \;=\; \frac{\sigma_{CK}\,\sigma_S^2 - \sigma_{SK}\,\sigma_{SC}}{\sigma_{SK}\,\sigma_C^2 - \sigma_{CK}\,\sigma_{SC}}$$

9.2 The optimum is a regression

Proposition 3. Let K = α + $\beta_S S + \beta_C C$ + e be the least-squares regression of the 1960 vote on the 1956 vote and the Catholic share, with $\beta_S > 0$. Then $k^* = \beta_C/\beta_S$, and $r(k^*)$ equals the multiple correlation coefficient R of that regression.

Proof. Correlation is invariant to positive rescaling, so corr(S + kC, K) = corr($\beta_S S + \beta_S kC$, K). Among all linear combinations of S and C, the fitted value $\beta_S S + \beta_C C$ has the highest correlation with K, and that correlation is R by definition. When $\beta_S > 0$, the fitted value is a positive multiple of S + (β_C/β_S) C, so the maximum of r(k) is attained at k = β_C/β_S and equals R. □

We confirmed this numerically. With our data the closed form gives $k^* = 0.5886$, the regression gives β_C/β_S = 0.39137 / 0.66493 = 0.5886, and both $r(k^*)$ and R equal 0.858.

9.3 Consequences

The lucky guess is not a single fortunate number. It is a two-predictor regression fitted to the very outcome it is scored against. Its .83 is an in-sample multiple correlation, while the simulation's .82 involved no parameter fitted to 1960. A fair comparison must therefore penalise the rule for fitting. Two simple penalties, computed with our placeholder data:

Measure	Value
In-sample r at k^* (= multiple R)	0.858
Adjusted R (32 states, 2 predictors)	0.847
Leave-one-out r (k^* re-estimated without each state)	0.841

The penalties are small here because the rule has only two free parameters and the fit is stable. The point stands nonetheless: the authors' comparison is between a forecast and a fit, and it favours the forecast more than the raw numbers suggest.

Proposition 3 also gives a check on the article. Because r(k) is smooth with a single maximum, the reported values constrain its curvature. A fall from .83 at k = 0.34 to below .80 at k = 0.32 implies a very narrow peak. In our placeholder data the peak is broad: r stays within 0.03 of its maximum for every k from 0.39 to 0.96. Whether the 1950s Catholic data reproduce the authors' sharp peak is a direct test of the printed table.

10. Replication results

The printed ranking holds up: it correlates 0.755 (Spearman) with Kennedy's actual 1960 share across the 32 states. Benchmarks that need only election returns reproduce closely. The one that needs 1950s Catholic data cannot yet be tested properly.

Benchmark	Pool and Abelson (1961)	This paper	Status
Simulation vs 1960 result	.82 (Pearson, unpublished index)	0.755 Spearman; 0.573 Kendall	Reproduced from printed ranking
Continuity with 1956	.39	0.427	Reproduced closely
Continuity with 1958	.37	—	Needs 1958 House vote by state
Continuity with 1956 and 1958	.44	—	Needs 1958 House vote by state
Contemporaneous trial heats	.53	—	Needs 1958 poll data
Best "lucky guess" k and r	0.34, .83	0.59, 0.858	Placeholder Catholic data only

10.1 The printed ranking

The article reports a product-moment correlation of .82 between its Kennedy index and the actual vote. The index itself is not printed, only the rank order it produced. We therefore correlate the ranks. The Spearman coefficient is 0.755 ($p = 5.9 \times 10^{-7}$), Kendall's tau 0.573, and the Pearson

correlation of rank with vote share 0.763. Using Kennedy's two-party share instead of his share of all votes changes the Spearman coefficient only to 0.763.

A rank correlation discards the spacing between states, which the Pearson coefficient on the index used. The two coefficients need not agree. Rhode Island and Massachusetts sit far above the other states in both the simulation and the result, and a Pearson coefficient on the index rewards that separation, which ranks ignore. A rank correlation of 0.755 is therefore compatible with the published .82, though it cannot confirm it.

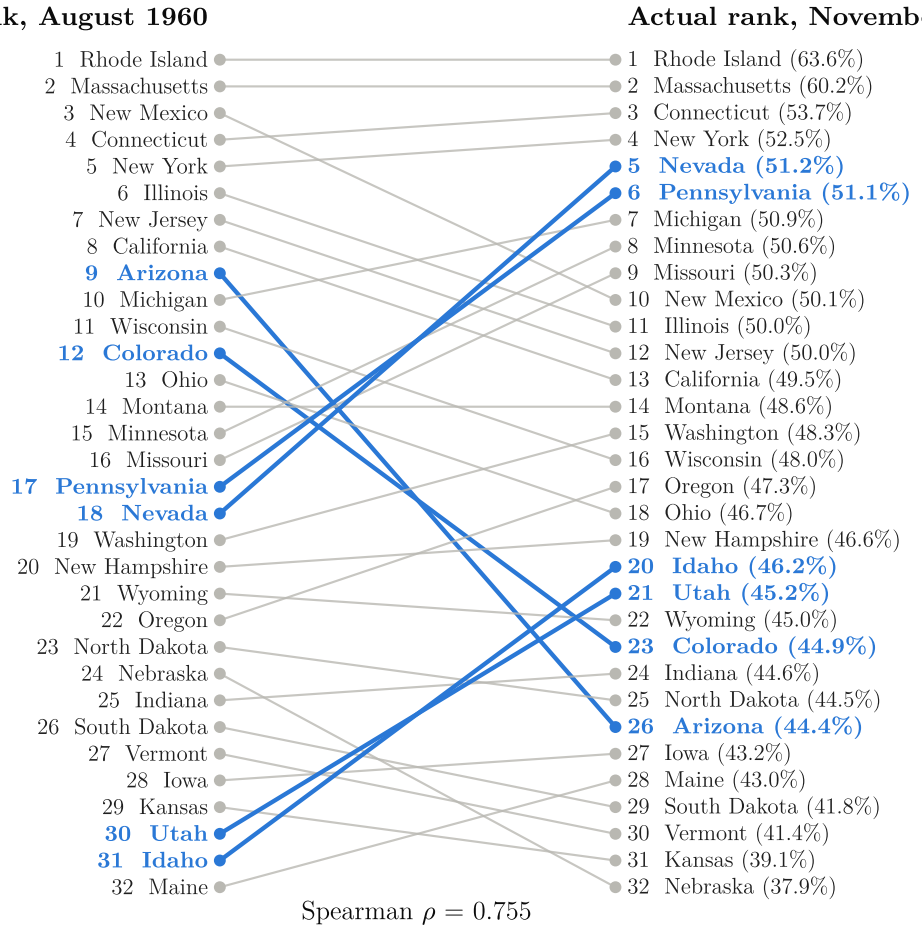


Figure 4: Simulmatics' ranking of the 32 states (August 1960, left) against Kennedy's actual rank by vote share (November 1960, right). Blue: the six states ranked nine or more places from where they finished; five are Western. Spearman $\rho = 0.755$. Sources: Pool and Abelson (1961, pp. 176–177); official 1960 returns. Animated version: Animation S4 in the online story map.

10.2 Where the model failed

The six largest rank errors are:

State	Region	Simulmatics rank	Actual rank	Error
Arizona	West	9	26	17 too high

State	Region	Simulmatics rank	Actual rank	Error
Nevada	West	18	5	13 too low
Pennsylvania	East	17	6	11 too low
Colorado	West	12	23	11 too high
Idaho	West	31	20	11 too low
Utah	West	30	21	9 too low

Four of the authors' five named failures (Arizona, Nevada, Idaho, Colorado) are among our six largest errors; New Mexico, the fifth, ranks 3 in the simulation and 10 in fact. Five of the six are in the West, consistent with the authors' diagnosis that pooling Mountain states with the Pacific coast broke the homogeneity assumption. The errors also run in both directions: Arizona and Colorado were overrated while Nevada, Idaho and Utah were underrated. A single regional bias would shift them all one way, so the failure is heterogeneity within the region rather than a regional offset. Pennsylvania is the one large Eastern miss and is not mentioned in the article.

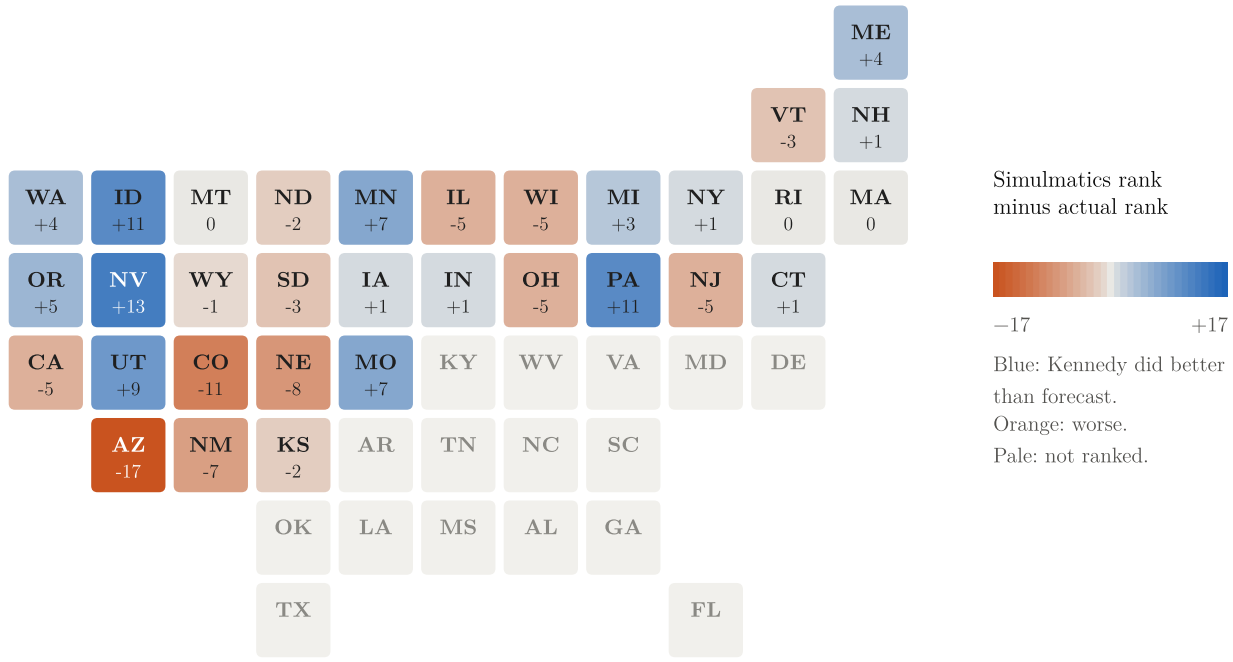


Figure 5: Simulmatics rank minus actual rank on a tile map (tile positions approximate geography). Blue: Kennedy did better than forecast (Nevada +13; Pennsylvania and Idaho +11; Utah +9). Orange: worse (Arizona −17; Colorado −11). The errors cluster in the West and run in both directions. Pale tiles: South and Border states, not ranked. Animated version: Animation S5 in the online story map.

10.3 The lucky guess with placeholder data

We could not obtain 1950s Catholic shares by state in machine-readable form for this draft. As a placeholder we use Gallup's pooled 2000–2004 self-identification estimates, labelled as such in

every output. With them the optimal k is 0.59, not 0.34, and the peak correlation is 0.858, not .83. The difference is expected. Catholic shares grew after 1960 in the Southwest through immigration, and self-identification counts differ from parish membership counts. The value of the exercise is the machinery, which will rerun unchanged on the proper series.

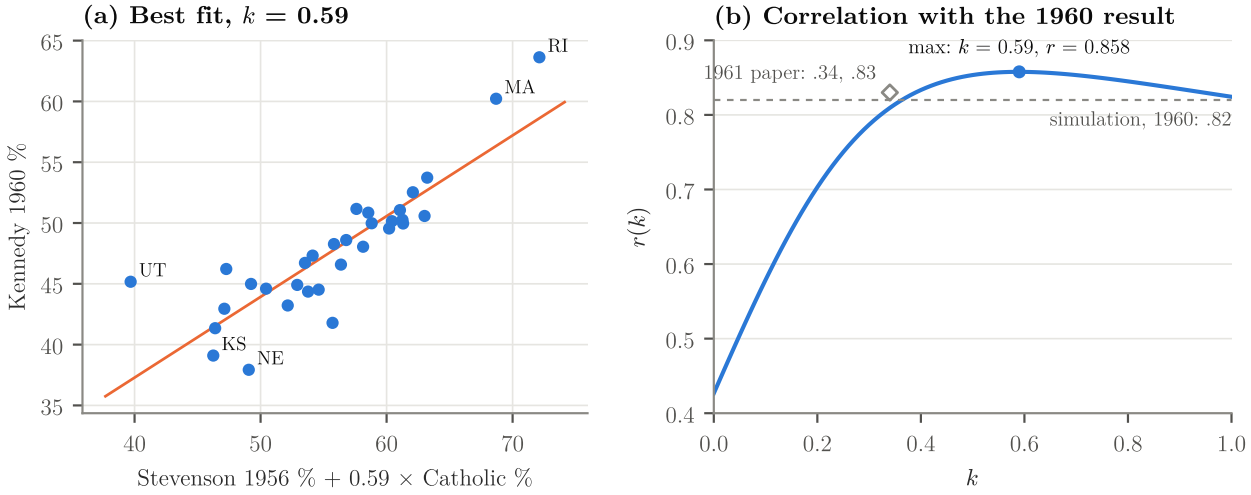


Figure 6: The lucky-guess benchmark with placeholder Catholic data (Gallup, 2000–2004). (a) 1956 Stevenson percentage plus 0.59× Catholic percentage against Kennedy's 1960 percentage. (b) The correlation $r(k)$ peaks at $k = 0.59$, $r = 0.858$; the authors reported $k = 0.34$, $r = .83$, against the simulation's .82 (Pool and Abelson 1961, p. 182). Animated version: Animation S6 in the online story map.

Some results do not depend on the placeholder. The swing from Stevenson in 1956 to Kennedy in 1960 averaged 8.3 points across the 32 states, from returns alone. Propositions 1–3 are mathematical and hold for any data. The swing's correlation with Catholic share, 0.81 with the placeholder, does depend on it and should be recomputed.

11. Animation and supplementary materials

The paper is accompanied by six animations (1920 × 1080, 30 frames per second, 14–20 seconds each), collected in an online story map, and a code kit that reproduces every number in Sections 7–10. Figures 1–6 show the final state of Animations S1–S6 in order: the typology (S1), the fourfold table (S2), the aggregation gap (S3), the shift from forecast to result (S4), the tile map of errors (S5) and the lucky-guess sweep (S6).

11.1 The lucky-guess animation (S6)

The animation visualises Section 9. Its left panel is a scatter of the 32 states: the 1960 Kennedy percentage on the vertical axis, and the predicted index S + kC on the horizontal. As k sweeps from 0 to 1, high-Catholic states slide right faster than the rest and the cloud straightens into a line. The right panel traces r(k) as it is computed, with the authors' optimum (k = 0.34, r = .83) and the simulation's .82 drawn as fixed references. After the sweep, k eases back to its optimum and holds.

The animation shows two things a static table cannot. It shows how much the ranking of states

changes as k moves, which is what a correlation summarises. And it shows the shape of the peak, which is the testable claim discussed at the end of Section 9.3. Every frame carries a footnote naming the placeholder Catholic series; rerunning the script with the 1950s series replaces both the frames and the label.

11.2 The code kit

File	Contents
simulmatics.py	The nine equation sets, the fourfold solver of Section 7 and the synthetic-state aggregator of Section 5
test_simulmatics.py	Twelve tests: margins and odds ratio recovered, independence at $\theta = 1$, each printed equation reproduced, monotonicity in P_{14} and θ, the full 480-type table run end to end
build_voter_types.py	Transcription of the 480 types with spot checks against the printed pages, and the state-to-region map
benchmarks.py	All benchmarks in Section 10, taking the Catholic series as an argument
render_sweep.py	The animation, taking the same argument

The kit is written so that missing inputs can be dropped in without code changes. The voter-type input table needs six columns per type (P_{56}, Q_{56}, P_{58}, Q_{58}, P_{35}, P_{14}) and the state weights three (state, type, persons).

12. Limitations and a research agenda

The mathematics can be recovered completely; the numbers cannot yet. Three inputs stand between this reconstruction and a full replication.

12.1 What is missing

The 480 × 52 matrix. A printout of the tape was given to the Roper Center as a condition of access (p. 171). Whether it survives is unknown to us. Failing that, the matrix can be rebuilt from the same archive: Gallup and other surveys of 1952–1958 that carry region, community size, sex, race, occupation, religion and party. Without the authors' question-to-cluster assignments and cut points, a rebuilt matrix will be a reconstruction, not a copy.

The state weights. These combine census counts by sex, community size, occupation and race with survey distributions of religion and party within each region. The weights were never released.

A modern version could use the IPUMS 1960 census sample for the census attributes and the rebuilt survey file for the rest.

Benchmark series. The 1952 enumeration of church membership by state from the National Council of Churches is the natural Catholic series; it is held by the Association of Religion Data Archives. The 1958 House vote by state is in the Clerk of the House's official statistics for that election.

12.2 Threats to validity

Even with all inputs, four sources of divergence remain. The cluster definitions involved unreported judgement (Section 4). The Black types need a religion assignment that the authors did not document (Section 3.4). The equations have unstated domain limits (Section 6.3). And the authors admit that clerical errors occurred under the two-week deadline, though none they found changed a conclusion (p. 173).

12.3 Agenda

1. **Sensitivity analysis.** Vary θ over [0.3, 1.5], the return fraction over [0.2, 0.5] and the turnout multipliers over [1, 3], and report the rank correlation over the grid. The authors called this "a full exploration" that "remains to be made" (p. 180).
2. **Homogeneity test.** Re-estimate with a separate Mountain region and measure how much of the Western error disappears.
3. **Shrinkage.** Replace complete pooling within region with partial pooling across states, turning the synthetic states into a multilevel model, and compare out-of-sample.
4. **The peak test.** Recompute r(k) with the 1952 Catholic series to see whether the authors' narrow optimum at 0.34 reappears (Section 9.3).
5. **Timing.** Compare the simulation with state-level trial heats month by month in 1960, which the authors could not do, to find when polls first became as predictive as the model.

13. Conclusion

The Simulmatics model of 1960 was a post-stratified, rule-based simulation with one estimated association and four judgemental constants. Its core is small enough to write on a page: nine pairs of equations, an odds-ratio constrained 2×2 table with a provably unique solution, and a weighted average over voter types.

The model's reported success survives the checks that public data allow. Its printed ranking of 32 states correlates 0.755 in rank with the outcome, and its known failures cluster where its own homogeneity assumption was weakest. The comparison the authors drew against a simple rule is fairer to them than it first appears, because that rule is a regression fitted after the fact.

What cannot be recovered from print is the data. The value of a reconstruction is that it states exactly what would be needed, so that the remaining steps are archival rather than conceptual.

Acknowledgements

I want to thank Jill Lepore for bringing to light the accomplishments and history of the Simulmatics Corporation and for sending me the archival files that contained much of the surviving Simulmatics code. I would also like to thank Bob Roswell, founder and director of the Systems Source Computer Museum, for many helpful discussions about the IBM hardware of the period. I also want to thank the librarians and archivists at MIT for their guidance in converting the .bin files that make up the Simulmatics code, in the Ithiel de Sola Pool archive. Special thanks also to Samuel Popkin for his comments on the use of Michigan Algorithm Decoder.

Data and code availability

All code, input tables and outputs are in the supplementary kit (simulmatics_kit.zip): the model, tests, voter-type table, state-region map, election returns, the printed ranking, benchmark outputs and the animation script. The placeholder Catholic series is in a file whose name marks it as a proxy. Every result in Section 10 regenerates with two commands once the 1952 series is supplied.

References

Berelson, B. R., Lazarsfeld, P. F. and McPhee, W. N. (1954). *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago: University of Chicago Press.

Burdick, E. (1964). *The 480*. New York: McGraw-Hill. Appendix, pp. 311–313.

Gelman, A. and Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology* 23(2), 127–135.

Lepore, J. (2020). *If Then: How the Simulmatics Corporation Invented the Future*. New York: Liveright.

McPhee, W. N. (undated). *A Model for Analyzing Macro-dynamics in Voting Systems*. Columbia University, Bureau of Applied Social Research.

Pool, I. de S. and Abelson, R. (1961). The Simulmatics Project. *Public Opinion Quarterly* 25(2), 167–183. JSTOR

Pool, I. de S., Abelson, R. P. and Popkin, S. L. (1964). *Candidates, Issues, and Strategies: A Computer Simulation of the 1960 Presidential Election*. Cambridge, MA: MIT Press.

Data sources

- 1960 United States presidential election, Wikipedia, results by state (accessed 21 November 2026).
- 1956 United States presidential election, Wikipedia, results by state (accessed 21 November 2026).
- Tracking Religious Affiliation, State by State, Gallup, pooled 2000–2004 (placeholder series).
- Churches and Church Membership in the United States, 1952 (States), Association of Religion Data Archives (intended Catholic series).