

SUPPLEMENTAL MATERIAL FOR
**Tests of Whether Candidate A or B Would Fare Better Against
C—and Other Polling Enrichment to Open Up Party Strategy**

The material below provides additional detail regarding different sections of the printed article, plus the Mathematical Appendix at the end.

The 2×2 Case

Table S1 offers a third example that has the same marginal totals as Examples 1 and 2 in the printed article. Example S1 has a smaller M and higher $|z|$, though. Formula (1) yields $z = (8 - 28)/\sqrt{8 + 28} = -20/6 = -3.33$, giving a significance level (two-tailed) of 0.0009. The corresponding exact probability is 0.0012.

Of course, the high values of $|z|$ in Examples 1 and S1 result from the lopsided split between x_{AC} and x_{CB} . For a given difference between the B-versus-C and A-versus-C responses (53% – 51%, or 2%, in this case), $|z|$ decreases with an increase in the sum M , the number of respondents who choose the Chartreuse candidate in one question and the Magenta opponent in the other. Comparison of Examples S1, 1, and 2 illustrates this point. The denominator of z is greatest in Example 2, intermediate in Example 1, and least in S1, but the numerator is the same (–20) in all three. Given a table with the set of four marginal totals of Table 1c, 1d, or S1 (530, 470; 510, 490), the numerator of z will be –20 regardless of how the rest of the table is filled in. In fact, the numerator has to be –20 for any set of marginal totals such that the sum of the first column exceeds the sum of the first row by 20.

With $z = -20/\sqrt{M}$ (the formula that thus applies with any 2% differential when $N = 1000$), one finds that A and B will be significantly different at the 0.05 level (two-tailed) so long as M , the number of respondents who divide their choices between the parties, is 94 or smaller. The value M

= 94 would represent a 37-to-57 split. (This result of 94 was actually found from exact probabilities, rather than from the normal approximation based on z .) With $z = -30/\sqrt{M}$ (corresponding to a 3% response difference between A and B vis-à-vis C, with $N = 1000$), significance is obtained if $M \leq 218$. The split would be 94 to 124 for $M = 218$. In various settings the great bulk of 1000 respondents may usually choose the same party on both questions so that a value above $M = 218$ may be uncommon—in which case a differential of 3% would generally be great enough to show firmly that B would be stronger than A vis-à-vis C. With $z = -10/\sqrt{M}$ (reflecting a 1% response difference), significance will result only if M does not exceed 20, a value whose split would be 5 to 15. Principles similar to those just covered for the 2×2 case apply also, though not with the same simplicity, to the 3×3 and 5×5 cases.

Table S1. Example S1

		B vs. C		
		B	C	(Total)
A vs. C	A	$(x_{AB} = 502)$	$x_{AC} = 8$	(510)
	C	$x_{CB} = 28$	$(x_{CC} = 462)$	(490)
(Total)		(530)	(470)	(1000)

$$N = 502 + 462 + 8 + 28 = 1000$$

$$M = 8 + 28 = 36$$

The 3 × 3 Case

To use formula (3) of the printed article, one has to choose a value for w . The choice will not matter as much in some situations as in others. One can calculate (3) with different values of w to see the impact of w . With the extreme of $w = 0$, the respondents who answer just one of the two questions are given no weight, which is equivalent to assuming that all such respondents are added (moved) to the group (x_{AB} or x_{CC}) in which the two question answers favor the candidate of the same party. If $w = 0$, (3) reduces to (1). Use of the other extreme, $w = 1$, is tantamount to assuming that all respondents who answer one question but not both are added to the group (x_{AC} or x_{CB}) in which the two question answers favor candidates of opposite parties. Since such respondents receive the same weight as respondents who answer both questions, the result is that (3) becomes the same as what (1) would be if x_{AC} , x_{A0} , and x_{0C} were combined into a single (x_{AC}) group and x_{CB} , x_{0B} , and x_{C0} were combined into a single (x_{CB}) group.

The middle value, $w = \frac{1}{2}$, is consistent with an assumption that all respondents who answer just one of the two questions will remain as refrainers on the other question and will not move (add) to either group to which they could go. It is also consistent with an assumption that these respondents will split evenly between moving to the group where the two favored candidates are of the same party and the group where they are of opposite parties. Most generally, use of $w = \frac{1}{2}$ is consistent with assuming that some of these respondents continue to refrain and the remainder split evenly between the two groups to which they could move. The choice $w = \frac{1}{2}$ would often seem to be a reasonable one.

To verify the statement that using $w = \frac{1}{2}$ is consistent with assuming that all refrainers remain as such, consider two (hypothetical) elections, A versus C and B versus C, and suppose that the N respondents are perfectly representative of the electorate. Suppose that any respondent who chooses neither candidate in one of the pairings will abstain in the associated election. Then the (sample) vote margin of A over C (negative if C is ahead) is

$$(x_{AB} + x_{A0} + x_{AC}) - (x_{CB} + x_{C0} + x_{CC}),$$

and of B over C is

$$(x_{AB} + x_{0B} + x_{CB}) - (x_{AC} + x_{0C} + x_{CC}).$$

The difference between these two vote margins (the first minus the second, representing the advantage of A over B in running against C) is

$$2(x_{AC} - x_{CB}) + (x_{A0} + x_{0C} - x_{0B} - x_{C0}) = 2D_0 + D_1,$$

which agrees precisely with twice the numerator of (3) when $w = \frac{1}{2}$.

Examples 3 (Table 2c), S2 (Table S2 below), and S3 (Table S3) illustrate (3) and the effect of w . In all three, though, a value of $\frac{1}{2}$ for w appears acceptable.

Example 3 may be rather typical, in the sense that D_1 has the same sign as D_0 and the ratio of $(x_{AC} + x_{CB})$ to M (the fraction of the M retained respondents who answer both questions) is fairly high (64 out of 100). Both of these characteristics can stabilize z as it relates to w . For Example 3, the value of z of (3) is relatively stable over the range of w (see Table S4), going from -2.25 at $w = 0$ to -2.341 at $w = \frac{1}{2}$ to -2.20 at $w = 1$. Actually, $|z|$ rises until it reaches its maximum at $w = \frac{32}{81}$ (where $z = -2.3467$) and then decreases from that point onward. Thus, z is easily significant at the 0.05 level (two-tailed) whatever the value of w .

In Example S2, a less favorable case, $(x_{AC} + x_{CB})$ is only half of M . The numerator of (3) does not change with w , because D_1 is 0, but the denominator does change (again see Table S4). Here, z of (3) varies more than in Example 3, going from -3.33 at $w = 0$ to -2.98 at $w = \frac{1}{2}$ to -2.36 at $w = 1$.

Example S3 was deliberately constructed as an apparently farfetched case to show how variation in z across different w could be bothersome. As Table S4 shows, z of (3) goes from -3.00 at $w = 0$ to -0.88 at $w = \frac{1}{2}$ to $+0.67$ at $w = 1$, with $z = 0$ at $w = 0.75$. The sign change of z results from D_1 being not only opposite in sign from D_0 but greater in absolute value as well.

Table S2. Example S2

		B vs. C				
		B	Neither	C	(Total)	
A vs. C	A	$(x_{AB} = 443)$	$x_{A0} = 9$	$x_{AC} = 8$	(460)	
	Neither	$x_{0B} = 9$	$(x_{00} = 82)$	$x_{0C} = 9$	(100)	
	C	$x_{CB} = 28$	$x_{C0} = 9$	$(x_{CC} = 403)$	(440)	
(Total)		(480)	(100)	(420)	(1000)	

$$N = 443 + 82 + 403 + 9 + 8 + 9 + 9 + 28 + 9 = 1000$$

$$M = 9 + 8 + 9 + 9 + 28 + 9 = 72$$

$$D_0 = 8 - 28 = -20$$

$$D_1 = 9 + 9 - 9 - 9 = 0$$

Table S3. Example S3

		B vs. C				
		B	Neither	C	(Total)	
A vs. C	A	$(x_{AB} = 385)$	$x_{A0} = 50$	$x_{AC} = 35$	(470)	
	Neither	$x_{0B} = 30$	$(x_{00} = 40)$	$x_{0C} = 30$	(100)	
	C	$x_{CB} = 65$	$x_{C0} = 10$	$(x_{CC} = 355)$	(430)	
(Total)		(480)	(100)	(420)	(1000)	

$$N = 385 + 40 + 355 + 50 + 35 + 30 + 30 + 65 + 10 = 1000$$

$$M = 50 + 35 + 30 + 30 + 65 + 10 = 220$$

$$D_0 = 35 - 65 = -30$$

$$D_1 = 50 + 30 - 30 - 10 = 40$$

Table S4. Numerator (Num.), denominator (Den.), and value of z of (3), for Examples 3, S2, and S3, for selected values of w

w	Example 3			Example S2			Example S3		
	Num.	Den.	z	Num.	Den.	z	Num.	Den.	z
0	-18	8.00	-2.25	-20	6.00	-3.33	-30	10.00	-3.00
0.25	-19	8.14	-2.33	-20	6.18	-3.23	-20	10.37	-1.93
0.5	-20	8.54	-2.34	-20	6.71	-2.98	-10	11.40	-0.88
0.75	-21	9.18	-2.29	-20	7.50	-2.67	0	12.94	0
1	-22	10.00	-2.20	-20	8.49	-2.36	10	14.83	+0.67

The 5 × 5 Case

The null hypothesis for the 5 × 5 case is H_0 of (S2) below. Table S5 below is analogous to Tables 1b and 2b of the printed article and shows the conditional probability that a respondent R is in a given cell in the table conditional on R not belonging to any of the five cells on the main diagonal. The sum of the 20 probabilities in Table S5 is, of course, equal to 1. Define

$$\begin{aligned}
 F_{A1} &= p_{Ac} + p_{aC}, F_{1B} = p_{cB} + p_{Cb}; \\
 F_{A2} &= p_{A0} + p_{ac} + p_{0C}, F_{2B} = p_{0B} + p_{cb} + p_{C0}; \text{ and} \\
 F_{A3} &= p_{Ab} + p_{a0} + p_{0c} + p_{cC}, F_{3B} = p_{aB} + p_{0b} + p_{c0} + p_{Cc}.
 \end{aligned}
 \tag{S1}$$

The equations (S1) are analogous to (4) of the printed article. As with (4) and Table 3, the three lines of (S1) relate to the cells in Table S5 that are, respectively, one, two, and three positions removed from the lower left or upper right corner of the table and thus reflect successively weaker differences between A and B. As a null hypothesis of no difference between A and B as to their strength against C, we specify

$$\begin{aligned}
 H_0: \quad & p_{AC} + w_1 F_{A1} + w_2 F_{A2} + w_3 F_{A3} = p_{CB} + w_1 F_{1B} + w_2 F_{2B} + w_3 F_{3B}, \\
 \text{or} \quad & (p_{AC} - p_{CB}) + w_1(F_{A1} - F_{1B}) + w_2(F_{A2} - F_{2B}) + w_3(F_{A3} - F_{3B}) = 0,
 \end{aligned}
 \tag{S2}$$

where (as indicated in the printed article) w_1, w_2, w_3 are to be chosen so that $1 \geq w_1 \geq w_2 \geq w_3 \geq 0$ (e.g., $w_1 = 3/4, w_2 = 1/2, w_3 = 1/4$) in order to reflect the successively decreasing informativeness.

Table S5. Conditional probabilities (5 × 5 case)

		B vs. C				
		B	B-leaner	Neither	C-leaner	C
A vs. C	A	----	p_{Ab}	p_{A0}	p_{Ac}	p_{AC}
	A-leaner	p_{aB}	----	p_{a0}	p_{ac}	p_{aC}
	Neither	p_{0B}	p_{0b}	----	p_{0c}	p_{0C}
	C-leaner	p_{cB}	p_{cb}	p_{c0}	----	p_{cC}
	C	p_{CB}	p_{Cb}	p_{C0}	p_{Cc}	----

Empirical illustrations

The citations for the data (poll questions) that we used to obtain the results shown in Tables 4, 5, and 6a of the printed article are as follows:

Cable News Network. CNN/ORC Polls. Opinion Research Corporation [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, *RoperExpress* [distributor]. Six polls (with respective access dates of 5/25/17, 5/25/17, 5/25/17, 5/19/17, 5/12/17, 5/10/17):
Jul, 2015 [dataset]. USORCCNN2015-007—columns 56-61 (for candidate questions).
Aug, 2015 [dataset]. USORCCNN2015-008—columns 84-87.
Sep, 2015 [dataset]. USORCCNN2015-009—columns 166-171.
Oct, 2015 [dataset]. USORCCNN2015-011—columns 90-101.
Nov, 2015 [dataset]. USORCCNN2015-012—columns 82-91.
Dec, 2015 [dataset]. USORCCNN2015-013—columns 52-57.

For the two polls used for Table 6b of the printed article, the respective data citations are:

CBS News and The New York Times. CBS News/New York Times Poll. Social Science Research Solutions (SSRS) [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, *RoperExpress* [distributor], accessed 8/31/17. Mar, 2016 [dataset]. USCBSNYT2016-0320, Version 2—columns 138-141.

CBS News. CBS News Poll. Social Science Research Solutions (SSRS) [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, *RoperExpress* [distributor], accessed 8/31/17. Apr, 2016 [dataset]. USCBS2016-0412A, Version 2—columns 212-217.

In both of Examples 4 and 5 (Tables 4 and 5), $|z|$ based on the weighted data is smaller than $|z|$ based on the unweighted data. That relation holds for most, though not all, of the triples in Tables 6a and 6b. Applying the weights always causes a relative increase in the denominator of $|z|$, the more so the greater the design effect. But the numerator can move in either direction, since the A-versus-B comparison (against C) can change to relatively benefit either A or B upon application of the weights.

The calculations of s and S are perhaps most easily explained by looking at the cell in Table 5 for the respondents who answer "Clinton" for the Clinton-vs.-Trump question and "Neither" for the Biden-vs.-Trump question. For the four respondents in that cell, the weights shown on the data

file are 0.95, 0.66, 0.93, 1.17 (it shows all weights to two decimal places). The sum of the weights across all 1012 respondents is 983.19. The calculations produce $(100/983.19)(0.95 + 0.66 + 0.93 + 1.17) = 0.38$ (percent) for s and $(100/983.19)^2(0.95^2 + 0.66^2 + 0.93^2 + 1.17^2) = 0.037$ for S , the values shown in Table 5.

Two details related to the 5×5 structure in the last three of the six polls in Table 6a need to be mentioned. First, the respondents who did not answer X or Y when asked to choose between them were then asked which way they leaned only if, for the first question, they were recorded as "Neither" or "Don't know/undecided/refused" as opposed to "Other." Second, for the four "Published results" columns in Table 6a, CNN/ORC combined X- (or Y-) leaners with those respondents who chose X (or Y) on the first question. Our calculations for the following column (which we are calling "Diff") were done on the same basis.

We can use Table 5 to illustrate the calculation of Diff (for the 3×3 case). The marginal totals for s show Clinton with 50.37% to 45.69% against Trump and Biden with 56.54% to 41.11% (which round, respectively, to 50 to 46 and 57 to 41, in agreement with the values shown for "Published results" for this triple in Table 6a). Then $\text{Diff} = \frac{1}{2}[(50.37 - 45.69) - (56.54 - 41.11)] = -5.38\%$, the value that Table 6a shows. Note that -5.38 is also the numerator of the z with weighting (6) for $w = \frac{1}{2}$, as given in Table 5.

In fact, for the 3×3 case, the numerator of z (6) with $w = \frac{1}{2}$ will always be identical to Diff. This fact may provide further argument in favor of choosing $w = \frac{1}{2}$ for the 3×3 case. For the 5×5 case, the numerator of z [from the analogue of (6)] will not generally be identical to Diff, because of different treatment of leaners.

Observe that weighting can affect not only the McNemar outcomes but also the basic poll results. Thus, in Table 5, the weighting boosts both Clinton and Biden vis-à-vis Trump. With the weights, Clinton even changes from being behind Trump to ahead of him.

For the values of both z and Diff in Tables 6a and 6b, a positive number means that A is stronger than B as an opponent of C and a negative number means that B is stronger. For the 3×3 case, z and Diff will always have the same sign. For the 5×5 case, it is mathematically possible

for them to differ in sign, although that does not occur for any of the triples in the last three of the six polls in Table 6a.

The design effects shown in the first column of Tables 6a and 6b pertain to each poll overall and are obtained through the formula $N\Sigma W^2/(\Sigma W)^2$, where W denotes the weight for an individual respondent and the summations are across all N respondents. The values in the column headed "Design effect" are for individual triples and are calculated with the formula $M\Sigma W^2/(\Sigma W)^2$, where the summations are over just the M off-diagonal respondents for the triple. Providing the design effects, which may seem rather high, gives some sense of the impact of the weights.

Further polling enrichment

Table S6 presents the raw data for Example 6 of the printed article, both for the 400 Magenta respondents and for the entire sample of 1000. The six one-on-one questions have $2^6 = 64$ possible combinations of responses, but only $4! = 24$ of them are transitive. Of those 24, only 11 appear in Table S6; it was best to construct an example of manageable size.

Table S7 shows how one obtains the values for the McNemar results (Table 7) from Table S6. Table S8 shows the conversion of the Magenta values in Table S6 to the data used in Table 8.

For the three-way (C, E, G and C, F, G) and four-way (C, E, F, G) general-election scenarios, we prescribe (again to keep the example manageable) that all 1000 respondents name their first preference when asked for whom they would vote in the usual plurality election. More realistically, of course, some voters would choose to vote for a less-preferred candidate to avoid "wasting" their votes. For the Magenta primary, though, the example does have 18 respondents who answer that they would vote for their second preference.

For each of the three multicandidate general-election scenarios (occasioned by the speculated independent candidacy of E or F, or both, if G were to win the Magenta primary), Tables S9 and S10 offer divergent results. Both tables are derived from Table S6.

Per Table S9, C is the plurality winner with 43% in all three cases. Second place goes to E in the first scenario, F in the second, and G in the third.

Table S10 shows the six one-on-one (Condorcet polling) outcomes. (Half of them repeat the columns in the lower part of Table S7; the other half, though, do not duplicate the right part of Table 8, because the latter pertains only to Magenta respondents.) The results in Table S10 (unlike those of Table S9) do not differ for the three scenarios, though the comparisons involving F and E are inapplicable for the first and second scenarios, respectively. C is the Condorcet loser in all three cases (though C is not far behind any of its three rivals). The Condorcet winners are E in the first and third scenarios and F in the second. G, the plurality winner for the Magenta primary with 44% of the choices (Table 8), has an especially poor one-on-one showing against each of E and F

among the full electorate, losing heavily by 75% to 25% in each comparison. E is favored over F by 59% to 41%.

Table S6. Raw data for Example 6

One-on-one question						Candidate ordering	No. of respondents		
C vs. E	C vs. F	C vs. G	E vs. F	E vs. G	F vs. G		Magenta voters	Other	Total
C	C	C	E	E	F	CEFG	0	276	276
C	C	C	F	E	F	CFEG	0	154	154
E	F	C	E	E	F	EFCG	14	6	20
E	F	G	E	E	F	EFGC	84	36	120
E	F	C	F	E	F	FECG	28	12	40
E	F	G	F	E	F	FEGC	98	42	140
C	C	G	E	G	G	GCEF	19	8	27
E	C	G	E	G	G	GECF	9	4	13
E	F	G	E	G	G	GEFC	94	40	134
C	F	G	F	G	G	GFCE	3	0	3
E	F	G	F	G	G	GFEC	51	28	73
						Total	400	600	1000

Note: Besides the questions above, all 1000 respondents were asked to pick a single candidate whom they would vote for (as in plurality voting) from among C, E, G; among C, F, G; and among C, E, F, G. For those three questions, all respondents picked their first-choice candidate. In addition, the 400 Magenta voters alone were asked for a single choice among E, F, G. All picked their first-choice candidate except for 12 of the 14 EFCG Magenta voters and 6 of the 84 EFGC Magenta voters, all 18 of whom named their second choice, F.

Table S7. Derivation of Table 7 of the printed article from Table S6 above

Candidate ordering	No. of respondents	Comparison, against C, of					
		E and F		E and G		F and G	
		E in E vs. C but C in F vs. C	C in E vs. C but F in F vs. C	E in E vs. C but C in G vs. C	C in E vs. C but G in G vs. C	F in F vs. C but C in G vs. C	C in F vs. C but G in G vs. C
		CEFG	276	0	0	0	0
CFEG	154	0	0	0	0	0	0
EFCG	20	0	0	20	0	20	0
EFGC	120	0	0	0	0	0	0
FECG	40	0	0	40	0	40	0
FEGC	140	0	0	0	0	0	0
GCEF	27	0	0	0	27	0	27
GECF	13	13	0	0	0	0	13
GEFC	134	0	0	0	0	0	0
GFCE	3	0	3	0	3	0	0
GFEC	73	0	0	0	0	0	0
Total	1000	13	3	60	30	60	40

(continued)

Candidate ordering	No. of respondents	Marginal results for					
		E vs. C		F vs. C		G vs. C	
		E	C	F	C	G	C
CEFG	276	0	276	0	276	0	276
CFEG	154	0	154	0	154	0	154
EFCG	20	20	0	20	0	0	20
EFGC	120	120	0	120	0	120	0
FECG	40	40	0	40	0	0	40
FEGC	140	140	0	140	0	140	0
GCEF	27	0	27	0	27	27	0
GECF	13	13	0	0	13	13	0
GEFC	134	134	0	134	0	134	0
GFCE	3	0	3	3	0	3	0
GFEC	73	73	0	73	0	73	0
Total	1000	540	460	530	470	510	490

Table S8. Derivation of the E, F, G candidate orderings in Table 8 of the printed article from the C, E, F, G candidate orderings in Table S6 above (Magenta respondents only)

C, E, F, G candidate ordering	No. of Magenta respondents	<u>E, F, G candidate ordering</u>			
		EFG	FEG	GEF	GFE
CEFG	0	0	0	0	0
CFEG	0	0	0	0	0
EF CG*	14	14	0	0	0
EF GC*	84	84	0	0	0
FECG	28	0	28	0	0
FEGC	98	0	98	0	0
GCEF	19	0	0	19	0
GE CF	9	0	0	9	0
GE FC	94	0	0	94	0
GFCE	3	0	0	0	3
GF EC	51	0	0	0	51
Total	400	98	126	122	54

*In regard to the first column of Table 8, see note under Table S6 concerning plurality vote choice.

Table S9. Plurality vote choices for the multicandidate general-election scenarios

Candidate ordering	No. of respondents	Plurality vote choice									
		Among C, E, G			Among C, F, G			Among C, E, F, G			
		C	E	G	C	F	G	C	E	F	G
CEFG	276	276	0	0	276	0	0	276	0	0	0
CFEG	154	154	0	0	154	0	0	154	0	0	0
EFCG	20	0	20	0	0	20	0	0	20	0	0
EFGC	120	0	120	0	0	120	0	0	120	0	0
FECG	40	0	40	0	0	40	0	0	0	40	0
FEGC	140	0	140	0	0	140	0	0	0	140	0
GCEF	27	0	0	27	0	0	27	0	0	0	27
GECF	13	0	0	13	0	0	13	0	0	0	13
GEFC	134	0	0	134	0	0	134	0	0	0	134
GFCE	3	0	0	3	0	0	3	0	0	0	3
GFEC	73	0	0	73	0	0	73	0	0	0	73
Total	1000	430	320	250	430	320	250	430	140	180	250

Table S10. One-on-one (Condorcet polling) results for the multicandidate general-election scenarios

Candidate ordering	No. of respondents	One-on-one question											
		C vs. E		C vs. F		C vs. G		E vs. F		E vs. G		F vs. G	
		C	E	C	F	C	G	E	F	E	G	F	G
CEFG	276	276	0	276	0	276	0	276	0	276	0	276	0
CFEG	154	154	0	154	0	154	0	0	154	154	0	154	0
EFCG	20	0	20	0	20	20	0	20	0	20	0	20	0
EFGC	120	0	120	0	120	0	120	120	0	120	0	120	0
FECG	40	0	40	0	40	40	0	0	40	40	0	40	0
FEGC	140	0	140	0	140	0	140	0	140	140	0	140	0
GCEF	27	27	0	27	0	0	27	27	0	0	27	0	27
GECF	13	0	13	13	0	0	13	13	0	0	13	0	13
GEFC	134	0	134	0	134	0	134	134	0	0	134	0	134
GFCE	3	3	0	0	3	0	3	0	3	0	3	0	3
GFEC	73	0	73	0	73	0	73	0	73	0	73	0	73
Total	1000	460	540	470	530	490	510	590	410	750	250	750	250

Mathematical Appendix

This appendix shows that, under the null hypothesis H_0 of (2) (in the printed article) for the 3×3 case [or (S2) above for the 5×5 case], the expression under the square-root sign in the denominator of (3) [or (5)] estimates the variance of its numerator. To start, let \mathbf{x} ($k \times 1$) = $(x_1, x_2, \dots, x_k)'$ be a sample of size M (where $x_1 + x_2 + \dots + x_k = M$) drawn from a k -variate multinomial distribution with \mathbf{p} ($k \times 1$) = $(p_1, p_2, \dots, p_k)'$ as its parameter vector (where $p_1 + p_2 + \dots + p_k = 1$). [For application to (3), $k = 6$.] The expectation vector of \mathbf{x} is, of course, $E(\mathbf{x}) = M\mathbf{p}$. Also well known (e.g., Lehmann, 1975: 380-381) is that $\text{cov}(p_g, p_G) = -Mp_g p_G$ for $g \neq G$ and $\text{var}(x_g) = Mp_g(1 - p_g)$. Thus, the $k \times k$ variance matrix for \mathbf{x} is

$$\text{var}(\mathbf{x}) = M\mathbf{D}_p - M\mathbf{p}\mathbf{p}' , \quad (\text{A1})$$

where \mathbf{D}_p denotes a $k \times k$ diagonal matrix with p_1, p_2, \dots, p_k along its main diagonal.

Consider a null hypothesis of the form

$$H_0: \mathbf{h}'\mathbf{p} = 0, \quad (\text{A2})$$

where \mathbf{h} ($k \times 1$) = $(h_1, h_2, \dots, h_k)'$ and at this point we leave the coefficients in \mathbf{h} unspecified. Then consider $\mathbf{h}'\mathbf{x}$. In general, $E(\mathbf{h}'\mathbf{x}) = \mathbf{h}'E(\mathbf{x}) = M\mathbf{h}'\mathbf{p}$, but this is 0 if H_0 (A2) holds. Using (A1), one finds that, for general \mathbf{h} , $\mathbf{h}'\mathbf{x}$ has variance

$$\text{var}(\mathbf{h}'\mathbf{x}) = \mathbf{h}'[\text{var}(\mathbf{x})]\mathbf{h} = M\mathbf{h}'\mathbf{D}_p\mathbf{h} - M\mathbf{h}'\mathbf{p}\mathbf{p}'\mathbf{h} = M\mathbf{h}'\mathbf{D}_p\mathbf{h} - M(\mathbf{h}'\mathbf{p})^2 . \quad (\text{A3})$$

If H_0 (A2) is true, though, then (A3) becomes just

$$\text{var}(\mathbf{h}'\mathbf{x}) = M\mathbf{h}'\mathbf{D}_p\mathbf{h} = M\sum_{g=1}^k h_g^2 p_g . \quad (\text{A4})$$

We now apply the results of the preceding paragraph to justifying (3). With $k = 6$, one sets $\mathbf{x} = (x_{AC}, x_{CB}, x_{A0}, x_{0C}, x_{0B}, x_{C0})'$, $\mathbf{p} = (p_{AC}, p_{CB}, p_{A0}, p_{0C}, p_{0B}, p_{C0})'$, and $\mathbf{h} = (1, -1, w, w, -w, -w)'$. [We remark that, more generally, one could set $\mathbf{h} = (1, -1, w_{A0}, w_{0C}, -w_{0B}, -w_{C0})'$ to allow differing weights for the four groups that answer just one of the two questions, but the associated complexity would not seem desirable.] With \mathbf{x} , \mathbf{p} , and \mathbf{h} as indicated, H_0 of (A2) is the same as H_0 of (2); $\mathbf{h}'\mathbf{x}$ is the same as the numerator of (3) (and has expectation of 0 under H_0); and, from (A4) along with the relation $p_{A0} + p_{0C} + p_{0B} + p_{C0} = 1 - p_{AC} - p_{CB}$, one finds that

$$\begin{aligned}\text{var}(\mathbf{h}'\mathbf{x}) &= M\sum_{g=1}^k h_g^2 p_g = M(p_{AC} + p_{CB}) + Mw^2(p_{A0} + p_{0C} + p_{0B} + p_{C0}) \\ &= (1 - w^2)M(p_{AC} + p_{CB}) + w^2M\end{aligned}\quad (\text{A5})$$

if H_0 is true. Finally, replacing p_{AC} and p_{CB} in (A5) with their respective estimates x_{AC}/M and x_{CB}/M leads to the expression under the square-root sign in the denominator of (3).

For the 5×5 case, $k = 20$. The justification for (5) is similar to that for (3) and again applies (A4).

On an ending note, we observe that, almost trivially, the development of (A1)-(A4) can be used to confirm the correctness of z of (1). With $k = 2$, set $\mathbf{x} = (x_{AC}, x_{CB})'$, $\mathbf{p} = (p_{AC}, p_{CB})'$, and $\mathbf{h} = (1, -1)'$. Then $\mathbf{h}'\mathbf{x} = x_{AC} - x_{CB}$ is the numerator of (1). It has expectation of 0 under the null hypothesis H_0 (which is $\mathbf{h}'\mathbf{p} = p_{AC} - p_{CB} = 0$, or $p_{AC} = p_{CB} = 1/2$). From (A4) one obtains $\text{var}(\mathbf{h}'\mathbf{x}) = M[(+1)^2 p_{AC} + (-1)^2 p_{CB}] = M(p_{AC} + p_{CB}) = M$, which conforms with the denominator of (1).

Technical remark: Besides that of (3) and (5), another generalization of the McNemar test does exist for cases beyond 2×2 (Bhapkar, 1966; Maxwell, 1970; Stuart, 1955) but is not applicable to our problem. Its null hypothesis (equality of corresponding row and column sums) is broader than ours—and can be false (leading to possible rejection) even if A and B are of equal strength against C.

References for Mathematical Appendix

Bhapkar VP (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. *Journal of the American Statistical Association* 61(1): 228-235. doi: 10.2307/2283057.

Lehmann EL (1975). *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day.

Maxwell AE (1970). Comparing the classification of subjects by two independent judges. *The British Journal of Psychiatry* 116(535): 651-655. doi: 10.1192/bjp.116.535.651.

Stuart A (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika* 42(3-4): 412-416. doi: 10.2307/2333387.