Supplement to Paolino, Philip. "Predicted Probabilities and Inference with Multinomial Logit"

A Articles in *APSR*, *AJPS*, and *JOP* that use Multinomial Logit in Analysis

A.1 Articles that report marginal effects and standard errors or confidence intervals as part of the primary analysis

1. Potter, Phillip B. K. and Matthew A. Baum. 2014. "Looking for Audience Costs in all the Wrong Places: Electoral Institutions, Media Access, and Democratic Constraint." *Journal of Politics*. 76(1): 167-181. Potter and Baum use multinomial logit for a robustness check and report multinomial logit coefficients and their standard errors and predicted probabilities and confidence intervals in Table 1 and Figures 1-5 in the online supplemental appendix.

2. Alvarez, R. Michael, Ines Levin, and Lucas Núñez. 2017. "The Four Faces of Political Participation in Argentina: Using Latent Class Analysis to Study Political Behavior." *Journal of Politics*. 77(4):1386-1402. Alvarez, et al. report multinomial logit coefficients and their credible intervals in Figure 3 and marginal effects and their credible intervals in Figures 4-6.

3. Hatemi, Peter and Zoltán Fazekas. 2018. "Narcissism and Political Orientations." *American Journal of Political Science*. 62(4):873-888. Hatemi and Fazekas report first differences of predicted probabilities from a multinomial logit model with confidence intervals in Figure 5.

4. Branton, Regina, Valerie Martinez-Ebers, Tony E. Carey, Jr., and Tetsuya Matsubayashi. 2015. "Social Protest and Policy Attitudes: The Case of the 2006 Immigrant Rallies." *American Journal of Political Science*. 59(2):390-402. Branton, et al. report predicted probabili-

ties and first differences with standard errors in Tables 1 and 2 and full results of multinomial logit coefficients, predicted probabilities, and first differences with standard errors in Tables B through I of the supplemental online appendix.

5. Smidt, Corwin D. 2017. "Polarization and the Decline of the American Floating Voter." *American Journal of Political Science*. 61(2):365-381. Smidt reports partial multinomial logit coefficients and their standard errors in Table 3, with full results reported in online supplement Table D1. Tables 3 (partial) and D.2 (full) report percentage changes in predicted probabilities with standard errors.

6. Kosmidis, Spyros. 2018. "International Constraints and Electoral Decisions: Does the Room to Maneuver Attenuate Economic Voting?" *American Journal of Political Science*. 62(3):519-534. Kosmidis reports multinomial logit coefficients and their standard errors in Tables 3-5 and the predicted marginal effects and confidence intervals of interest in Figures 2-4.

7. Lerman, Amy E, Meredith L. Sadin, and Samuel Trachtman. 2017. "Policy Uptake as Political Behavior: Evidence from the Affordable Care Act." *American Political Science Review*. 111(4):755-770. Lerman, et al. report marginal effects for predicted probabilities and their confidence intervals in Figure 2.

8. Hale, Henry E. and Timothy J. Colton. 2017. "Who Defects? Unpacking a Defection Cascade from Russia's Dominant Party 2008-12." *American Political Science Review*. 111(2):322-337. Hale and Colton report first differences for predicted probabilities with confidence intervals in Table 5.

9. Gans-Morse, Jordan. 2017. "Demand for Law and the Security of Property Rights: The Case of Post-Soviet Russia." *American Political Science Review*. 111(2):338-359. Gans-Morse reports first differences of predicted probabilities with standard errors in Table 5.6.2 of the supplemental online appendix.

10. Eggers, Andrew C. and Arthur Spirling. 2014. "Ministerial Responsiveness in Westminster Systems: Institutional Choices and House of Commons Debates, 1832-1915." *American Journal of Political Science*. 58(4):873-887. Eggers and Spirling present boxplots of the predicted probabilities.

A.1.1 Articles that report marginal effects and standard errors or confidence intervals as part of the secondary analysis

1. Fortunato, David, Randolph T. Stevenson, and Greg Vonnahme. 2016. "Context and Political Knowledge: Explaining Cross-National Variation in Partisan Left-Right Knowledge." *Journal of Politics*. 78(4):1211-1228. Fortunato, Stevenson, and Vonnahme use multinomial logit to generate predicted probabilities that are then used as the dependent variables in second-stage compositional analysis.

A.2 Articles that do not report marginal effects and standard errors or confidence intervals

1. Utych, Stephen M. and Cindy D. Kam. 2014. "Viability, Information Seeking, and Vote Choice." *Journal of Politics*. 76(1):152-166. Utych and Kam report multinomial logit coefficients and the standard errors in Table 2 and predicted probabilities in the text, but without standard errors or confidence intervals for the predicted marginal effects.

2. Burden, Barry C. and Amber Wichowsky. 2014. "Economic Discontent as a Mobilizer: Unemployment and Voter Turnout." *Journal of Politics*. 76(4):887-898. Burden and Wichowsky report multinomial logit coefficients and their standard errors in Table 5 and predicted marginal effects in the text, but without standard errors or confidence intervals for the marginal effects. They write in an interpretation of their results in Table 5, "The results do show that higher state unemployment rates do in fact stimulate grater turnout (i.e. less ab-

stention) but that the increase in turnout only benefits Democratic gubernatorial candidates." Analysis of changes in the predicted probabilities indicates that Democratic vote does increase by an amount that meets conventional levels of statistical significance, but that the predicted decrease in abstention does not meet conventional levels, $p < .20$.

3. McCann, James A. and Katsuo A. Nishikawa Chávez. 2016. "Partisanship by Invitation: Immigrants Respond to Political Campaigns." *Journal of Politics*. 78(4):1196-1210. McCann and Nishikawa report multinomial logit coefficients in Table A2 as supplementary analysis in an online appendix to Table 2, but without predicted probabilities or standard errors or confidence intervals for the marginal effects.

4. Nall, Clayton. 2015. "The Political Consequences of Spatial Policies: How Interstate Highways Facilitated Geographic Polarization." *Journal of Politics*. 77(1):394-406. Nall reports multinomial logit coefficients as supplementary analysis in Table B.2 of an online appendix and ternary plots of the predicted probabilities in the text.

5. Chiba, Daina, Jesse C. Johnson, and Brett Ashley Leeds. 2015. "Careful Commitments: Democratic States and Alliance Design." *Journal of Politics*. 77(4):968-982. Chiba, Johnson, and Leeds report multinomial logit coefficients in Tables A.24 and A.25 as supplementary analysis in an online appendix, but without predicted probabilities or standard errors or confidence intervals for the marginal effects. An interesting note with the presentation of the coefficients in Table A.24 is that the authors run two sets of estimates, with a change in the baseline as the only difference.

6. Carnes, Nicholas and Meredith L. Sadin. 2015. "The 'Mill Worker's Son' Heuristic: How Voters Perceive Politicians from Working-Class Families–and How They Really Behave in Office." *Journal of Politics*. 77(1):285-298. Carnes and Sadin report multinomial logit coefficients as supplementary analysis in an online appendix, but without predicted probabilities or standard errors or confidence intervals for the marginal effects. The purpose of this analy-

sis is to check for balance across experimental subjects' characteristics assigned to different treatments

7. Kuo, Alexander, Neil Malhotra, and Cynthia Hyungjung Mo. 2017 "Social Exclusion and Political Identity: The Case of Asian American Partisanship." *Journal of Politics*. 79(1):17-32. Kuo, Malhotra, and Mo report multinomial logit coefficients and their standard errors as supplementary analysis in Table B.2 of an online appendix, but without predicted probabilities or standard errors or confidence intervals for the marginal effects. The results from the multinomial logit are consistent with the estimated marginal effects that are reported in the text, albeit also without standard errors or confidence intervals.

8. Curry, Todd A. and Mark S. Hurwitz. 2016. "Strategic Retirements of Elected and Appointed Justices: A Hazard Model Approach." *Journal of Politics*. 78(4):1061-1075. Curry and Hurwitz report multinomial logit coefficients and their standard errors in Table 4 and predicted probabilities in Figures 2 and 3, but without standard errors or confidence intervals.

9. Hajnal, Zoltan and Michael U. Rivera. 2014. "Immigration, Latinos, and White Partisan Politics: The New Democratic Defection." *American Journal of Political Science*. 58(4):773-789 Hajnal and Rivera report multinomial logit coefficients and their standard errors in Table 3 and selected predicted probabilities without standard errors or confidence intervals in the text.

10. Weber, Christopher R., Howard Lavine, Leonie Huddy, and Christoper M. Federico. 2014. "Placing Racial Stereotypes in Context: Social Desirability and the Politics of Racial Hostility." *American Journal of Political Science*. 58(1):63-78. Weber, et al. report multinomial logit coefficients and their standard errors in Tables 3 and 4 and predicted probabilities and confidence intervals for one option in Figure 3, but only predicted probabilities for the three other options. This presentation allows for an evaluation of one hypothesis, but not for

another claim in the text concerning one of the alternatives, providing midpoint responses (p.73). They also report multinomial logit coefficients in Tables A3, A4, A6, A7, A9, and A10 in the online supplemental appendix, but without predicted probabilities.

11. Burden, Barry C., David T. Canon, Kenneth R. Mayer, and Donald P. Moynihan. 2014. "Election Laws, Mobilization, and Turnout: The Unanticipated Consequences of Election Reform." *American Journal of Political Science*. 58(1):95-109. Burden et al. report multinomial logit coefficients and their standard errors in Table 2, but not predicted probabilities or marginal effects.

12. Lacina, Bethany. 2014. "How Governments Shape the Risk of Civil Violence: India's Federal Reorganization." *American Journal of Political Science*. 58(3):720-738. Lacina reports multinomial logit coefficients and their standard errors in Table 5 in the text and Table 5 in the online supplemental appendix and predicted probabilities, but without standard errors or confidence intervals in Figure 2 in the text and Figures 4 and 5 in the online supplemental appendix.

13. Laver, Michael and Kenneth Benoit. 2015. "The Basic Arithmetic of Legislative Decisions." *American Journal of Political Science*. 59(2):275-291. Laver and Benoit report multinomial logit coefficients (risk ratios) and their standard errors in Table 3, but without predicted probabilities.

14. Thachil, Tariq. 2014. "Elite Parties and Poor Voters: Theory and Evidence from India." *American Political Science Review*. 108(2):454-477. Thachil reports multinomial logit coefficients and their standard errors in Table 5, but without predicted probabilities or marginal effects.

A.2.1 Articles that do not report marginal effects and standard errors or confidence intervals as part of the secondary analysis

1. Galkwad, Nikhar and Gareth Nellis. 2017. "The Majority-Minority Divide in Attitudes toward Internal Migration: Evidence from Mumbai." *American Journal of Political Science*. 61(2):456-472. Galkwad and Nellis report multinomial logit coefficients in Table 2 as a test of covariate balance for their experiments.

B Articles in *IO*, *ISQ*, and *JCR* that use Multinomial Logit in Analysis

B.1 Articles that report marginal effects and standard errors or confidence intervals as part of primary analysis

1. Klein, Graig R. and Patrick M. Regan. 2018. "Dynamics of Political Protests." *International Organization*. 72(2):485-521. Klein and Regan report multinomial logit coefficients and their standard errors in Tables 2-4 and predicted probabilities with confidence intervals in Figures 8 and 9.

2. Walter, Stefanie, Elias Dinas, Ignacio Jurado, and Nikitas Konstantinidis. 2018. "Noncooperation by Popular Vote: Expectations, Foreign Intervention, and the Vote in the 2015 Greek Bailout Referendum." *International Organization*. 72(4):969-994. Walter, et al. report changes in predicted probabilities and their confidence intervals in Figure 8. They also report the multinomial logit coefficients and their standard errors in Table A.3 of the online supplemental appendix.

3. Gray. 2018. "Life, Death, or Zombie? The Vitality of International Organizations." *International Studies Quarterly*. 62(1):1-13. Gray presents multinomial logit coefficients and their standard errors in Table 2 and predicted probabilities with confidence intervals in Figure 2.

4. DiCicco, Jonathan M. and Benjamin O. Fordham. 2018. "The Things They Carried: Generational Effects of the Vietnam War on Elite Opinion." *International Studies Quarterly*. 62(1):131-144. DiCicco and Fordham present predicted probabilities and confidence intervals in Figure 1 of the text and report multinomial logit coefficients and their standard errors in Table A2 of the online supplemental appendix.

5. Sudduth, Jun Koga and Curtis Bell. 2018. "The Rise Predicts the Fall: How the Method of Leader Entry Affects the Methods of Leader Removal in Dictatorships." *International Studies Quarterly*. 62(1):145-159. Sudduth and Bell report multinomial logit coefficients and their standard errors in Tables 4 and 5 and marginal effects and their confidence intervals in Figures 1-3.

6. Matanock, Aila M.. 2018. "External Engagement: Explaining the Spread of Electoral Participation Provisions in Civil Conflict Settlements." *International Studies Quarterly*. 62(3):656-670. Matanock reports partial multinomial logit coefficients and their standard errors in Table 3 (with full coefficients in Tables A4.0-A4.3 and A8 in the online supplemental appendix) and predicted probabilities with confidence intervals in Figure 3.

7. Beardsley, Kyle and Nigel Lo. 2014. "Third-Party Conflict Management and the Willingness to Make Concessions." *Journal of Conflict Resolution*. 58(2):363-392. Beardsley and Lo report multinomial logit coefficients and their standard errors in Table 4 and predicted probabilities with confidence intervals in Figure 1.

8. Clare, Joe. 2014. "Hawks, Doves, and International Cooperation." *Journal of Conflict Resolution*. 58(7):1311-1337. Clare reports multinomial logit coefficients and their standard errors in Table 1 and predicted probabilities with confidence intervals in Figure 2.

9. Jaeger, David A, Esteban F. Klor, Sami H. Miaari, and M. Daniele Paserman. 2015. "Can Militants Use Violence to Win Public Support? Evidence from the Second Intifada." *Journal of Conflict Resolution*. 59(3):528-549. Jaeger, et al. report marginal effects with standard errors in Tables 2 and 3.

10. Oppenheim, Ben, Abbey Steele, Juan F. Vargas, and Michael Weintraub. 2015. "True Believers, Deserters, and Traitors: Who Leaves Insurgent Groups and Why." *Journal of Conflict Resolution*. 59(5):794-823. Oppenheim, et al. report multinomial logit coefficients and their standard errors in Tables 2 and 3 and marginal effects and confidence intervals in Figures 3 and 4. Some of the inferences, however, are based upon coefficients.

11. Butcher, Charles and Isak Svensson. 2016. "Manufacturing Dissent: Modernization and the Onset of Major Nonviolent Resistance." *Journal of Conflict Resolution*. 60(2):311-339. Butcher and Svensson report multinomial logit coefficients and their standard errors in Tables 1 and 2 and marginal effects and confidence intervals in Figures 2 and 3.

B.1.1 Articles that report marginal effects and standard errors or confidence intervals as part of the secondary analysis

1. Reid, Lindsay. 2017. "Finding a Peace that Lasts: Mediator Leverage and the Durable Resolution of Civil Wars." *Journal of Conflict Resolution*. 61(7):1401-1431. Reid reports multinomial probit coefficients in Table 1 and predicted probabilities with confidence intervals in Figure 1. Multinomial logit coefficients are used as a robustness check and are reported in Table 3 of the online supplemental appendix. Reid does not present predicted probabilities from the multinomial logit coefficients, but a reasonable supposition is that such data would have been reported if it had been used for the primary analysis.

B.2 Articles that do not report marginal effects and standard errors or confidence intervals as part of the primary analysis

1. Ahlquist, John S., Amanda B. Clayton, and Margaret Levi. 2014. "Provoking Preferences: Unionization, Trade Policy, and the ILWU Puzzle." *International Organization*. 68(1):33-75. Ahlquist, et al. present multinomial logit coefficients and their standard errors in Table 2 and predicted probabilities in ternary plots in Figures 3 and 4. An interesting note with the presentation of the coefficients in Table 2 is that the authors report an additional set of estimates, with a change in the baseline as the only difference.

2. Horowitz, Michael C and Allan C. Stam. 2014. "How Prior Military Experience Influences the Future Militarized Behavior of Leaders." *International Organization*. 68(3):527-559. Horowitz and Stam report multinomial logit coefficients and their standard errors, without predicted probabilities, in Table 7 of the supplemental online appendix as a robustness check.

3. Fortna, Virginia Page. 2015. "Do Terrorists Win? Rebels' Use of Terrorism and Civil War Outcomes." *International Organization*. 69(3):519-556. Fortna reports multinomial logit coefficients and their standard errors in Table 3, 5 and 6. Tables 4 and 7 present predicted probabilities and the significance levels of their differences, but without standard errors or confidence intervals for those differences.

4. Powell, Emilia Justyna. 2015. "Islamic Law States and Peaceful Resolution of Territorial Disputes." *International Organization*. 69(4):777-807. Powell reports multinomial logit coefficients and their standard errors in Table 3 and predicted probabilities without standard errors or confidence intervals in Table 4.

5. Buhaug, Halvard, Lars-Erik Cederman, and Kristian Skrede Gleditsch. 2014. "Square Pegs in Round Holes: Inequalities, Grievances, and Civil War." *International Studies Quarterly*. 58(2):418-431. Buhaug, et al. report multinomial logit coefficients and their standard errors

in Tables 1 and 3, but present selected predicted probabilities without standard errors or confidence intervals.

6. Albertus, Michael and Victor Menaldo. 2014. "Dealing with Dictators: Negotiated Democratization and the Fate of Outgoing Autocrats." *International Studies Quarterly*. 58(3):550-565. Albertus and Menaldo report multinomial logit coefficients and their standard errors in Table 5. The text reports predicted probabilities without standard errors for the model in columns 2a and 2b of Table 5 and predicted probabilities with significance levels only for columns 6a and 6b from Table 5.

7. Choi, Hyun Jin and Clionadh Raleigh. 2015. "Dominant Forms of Conflict in Changing Political Systems." *International Studies Quarterly*. 59(1):158-171. Choi and Raleigh report multinomial logit coefficients and their standard errors in Table 3, but without predicted probabilities.

8. Guisinger, Alexandra and Elizabeth N. Saunders. 2017. "Mapping the Boundaries of Elite Cues: How Elites Shape Mass Opinion across International Issues." *International Studies Quarterly*. 61(2):425-441. Guisinger and Saunders report coefficients and significance levels, but not standard errors, in Figure 2. They also report multinomial logit coefficients and standard errors in Table 2 and predicted first differences and significance levels, but not standard errors or confidence intervals in Table 3. It appears in Table 3 that the significance levels use standard errors of marginal effects, but this is not entirely clear.

9. Greenhill, Kelly M. and Ben Oppenheim. 2017. "Rumor Has It: The Adoption of Unverified Information in Conflict Zones." *International Studies Quarterly*. 61(3):660-676. Greenhill and Oppenheim report multinomial logit coefficients and their standard errors in Tables 4-7, but do no report predicted probabilities.

10. Hummel, Sarah. 2017. "Relative Water Scarcity and Country Relations along Cross-Boundary Rivers: Evidence from the Aral Sea Basin." *International Studies Quarterly*. 61(4):795-808.

Hummel reports multinomial logit coefficients and their standard errors in Table 2, but do not report predicted probabilities.

11. Kastner, Scott L. 2016. "Buying Influence? Assessing the Political Effects of China's International Trade." *Journal of Conflict Resolution*. 60(6):980-1007. Kastner reports multinomial logit coefficients and their standard errors in Tables 2, 4, and 6 reports predicted probabilities, but without standard errors or confidence intervals in the text.

12. Foster, Dennis M. 2017. "Inter Arma Silent Leges? Democracy, Domestic Terrorism, and Diversion." *Journal of Conflict Resolution*. 61(7):1371-1400. Foster reports multinomial logit coefficients and their standard errors in Table 3 and ratios of predicted probabilities, with significance levels, but without standard errors or confidence intervals in Table 4.

13. Bauer, Vincent, Keven Ruby, and Robert Pape. 2017. "Solving the Problem of Unattributed Political Violence." *Journal of Conflict Resolution*. 61(7):1537-1564. Bauer, et al. report multinomial logit coefficients (with significance levels, but not standard errors) in Table 4.

14. Gelpi, Christopher. 2017. "The Surprising Robustness of Surprising Events: A Response to a Critique of 'Performing on Cue "'*Journal of Conflict Resolution*. 61(8):1816-1834. Gelpi reports multinomial logit coefficients in Tables 1-3 and predicted probabilities without standard errors or confidence intervals in Figures 1-3.

15. Crisman-Cox, Casey. 2018. "Enemies within: Interactions between Terrorists and Democracies." *Journal of Conflict Resolution*. 62(8):1661-1685. Crisman-Cox reports multinomial logit coefficients in Table 3 and reports relevant marginal effects, but without standard errors or confidence intervals. There may not be much reason to report more, but since the full multinomial logits take up one page, one could argue that a more reasonable approach to presentation would be to present marginal effects statistics or plots for the relevant covariates.

B.2.1 Articles that do not report marginal effects and standard errors or confidence intervals as part of the secondary analysis

1. Chaudoin, Stephen. 2014. "Promises or Policies? An Experimental Analysis of International Agreements and Audience Reactions." *International Organization*. 68(1):235-256. Chaudoin uses multinomial logit as a check on ordered logit results that check for balance in experimental treatments, but does not present analysis in the text or online supplemental appendix other than to write that treatment was not related to free trade attitudes (p.246).

2. McManus, Roseanne W. and Keren Yarhi-Milo. 2017. "The Logic of 'Offstage' Signaling: Domestic Politics, Regime Type, and Major Power-Protégé Relations." *International Organization*. 71(4):701-733. McManus and Yarhi-Milo report multinomial logit coefficients and their standard errors without predicted probabilities in Tables A4 and A5 of the online supplemental appendix as a robustness check.

3. Prorok, Alyssa K. 2017. "The (In)compatibility of Peace and Justice? The International Criminal Court and Civil Conflict Termination." *International Organization*. 71(2)213-243. Prorok reports multinomial logit coefficients and their standard errors, but not predicted probabilities, as a robustness check in Table C of the online supplemental appendix.

4. Baccini, Leonardo, Andreas Dür, and Manfred Elsig. 2015. "The Politics of Trade Agreement Design: Revisiting the Depth–Flexibility Nexus." *International Studies Quarterly*. 59(4):765-775. Baccini, et al. write in footnote 15 that they use multinomial logit as a robustness check for ordered probit, but that results are available upon request. No further information is available in the replication files.

5. Avdan. 2014. "Controlling Access to Territory: Economic Interdependence, Transnational Terrorism, and Visa Policies." *Journal of Conflict Resolution*. 58(4):592-624. Avdan reports multinomial logit coefficients in Table 3 as a robustness check and some predicted probabilities without standard errors or confidence intervals in the text.

6. Daxecher, Ursula. 2017. "Dirty Hands: Government Torture and Terrorism." *Journal of Conflict Resolution*. 61(6):1261-1289. Daxecher reports multinomial logit coefficients in Table 2 as a way of addressing potential endogeneity, but without predicted probabilities, marginal effects, or their standard errors.

7. Balcells, Laia and Stathis N. Kalyvas. 2014. "Does Warfare Matter? Severity, Duration, and Outcomes of Civil Wars." *Journal of Conflict Resolution*. 58(8):1343-1359. Balcells and Kalyvas report multinomial logit coefficients, but not predicted probabilities or marginal effects in Tables A3, A5, A6, and A7 in the online supplemental appendix.

8. Aytac, S. Erdem, Luis Schiumerini, and Susan Stokes. 2018. "Why Do People Join Backlash Protests? Lessons from Turkey." *Journal of Conflict Resolution*. 62(6): 1205-1228. Aytac, et al. report a Wald statistic in a footnote and refer to results in Table A1 of the online supplemental appendix as a check on balancing between treatments; although, it appears that the figures may simply be the means of the covariates.

C An Example of a Baseline with Few Observations

An example of the problems with using MNL coefficients for inference comes from Gelpi (2017). This example uses people's perceptions of the chances of success in Iraq as the dependent variable from his study. In Table A.1, the analysis that Gelpi presents in his Table 2 for respondents who are very strong supporters of President Bush is replicated twice with two different baselines from the one he used. In the top half of Table A.1, a baseline category with very few observations is used, and the standard errors of the MNL coefficients for the treatment covariates are all in the thousands. It is easy to see that using these coefficients as a basis for inference would be very misleading.

Merely changing the baseline to a different category, "somewhat likely to succeed," produces standard errors in the category with very few observations remain extremely high, but with one exception, the other standard errors are now much smaller. In addition, the MNL coefficient for positive events is now statistically significant at $p < .10$, whereas the results that Gelpi reports in Table 2, using "not very likely" as a baseline is not significant. From his analysis, Gelpi argues, "neither positive news events nor confident messages from the President had any impact on the surge or success in Iraq among those who approve of Bush" (1832). This conclusion, however, is solely a product of the choice of a baseline.

As demonstrated in the text, the MNL coefficients still do not provide enough information to evaluate the effect of positive events upon these respondents' attitudes about the likelihood of success in Iraq. Holding other treatments at 0, exposure to positive events has a .227 (se=.103) increase in the probability of respondents saying that the likelihood of success is "very likely" is now statistically significant at $p < .05$ and larger than the effects of positive events on respondents who "strongly disapprove" and "somewhat disapprove" of President Bush (Table A.2). The results in Table A.2 suggest an interpretation of positive events in this case as one where exposure to positive events led most respondents to have a more positive assessment of the US's chances of success in Iraq.

Table A.1: Multinomial Logit with Different Baselines

	Expectations of Success in Iraq			
	Baseline Category = Not at all likely			
	Not at all likely	Not very likely	Somewhat likely	Very likely
Positive Events	0.000	15.727	15.141	16.183
	(.)	(3839.373)	(3839.373)	(3839.373)
Negative Events	0.000	16.464	15.304	16.068
	(.)	(3611.992)	(3611.992)	(3611.992)
Confident Bush	0.000	−0.966	14.192	14.734
	(.)	(4851.669)	(3898.236)	(3898.236)
Negative Bush	0.000	16.541	15.633	15.182
	(.)	(3146.423)	(3146.422)	(3146.422)
Constant	0.000	−0.933	2.098	2.599
	(.)	(1.596)	(1.051)	(1.033)
	Baseline Category = Somewhat likely			
	Not at all likely	Not very likely	Somewhat likely	Very likely
Positive Events	−15.141	0.585	0.000	1.042
	(3839.373)	(1.519)	(.)	(0.567)
Negative Events	−15.304	1.160	0.000	0.764
	(3611.992)	(1.327)	(.)	(0.541)
Confident Bush	−14.192	−15.158	0.000	0.543
	(3898.236)	(2888.330)	(.)	(0.667)
Cautious Bush	−15.633	0.908	0.000	−0.451
	(3146.423)	(1.248)	(.)	(0.497)
Constant	−2.098	−3.031	0.000	0.501
	(1.051)	(1.273)	(.)	(0.392)
N	118			

Entries are maximum likelihood coefficients with standard errors in parentheses for respondents who strongly approve of President Bush.

Table A.2: Positive Events and Expectations of Success in Iraq

	Approval of President Bush			
Success:	Strongly Disapprove	Somewhat Disapprove	Somewhat Approve	Strongly Approve
Not at all likely	-.008 (.059)	.002 (.041)	.000 (.000)	-.043 (.043)
Not very likely	-.094 (.060)	-.280 (.101)	-.012 (.033)	-.002 (.023)
Somewhat likely	.101 (.045)	.187 (.107)	.064 (.073)	-.181 (.100)
Very likely	-.000 (.000)	.091 (.071)	-.052 (.072)	.227 (.103)
Observations	449	145	284	118

Entries are the change in the predicted probability of each response (and standard error) for a respondent receiving the "positive events" treatment compared with a respondent who did not receive any treatment.

D Using MNL Interaction Coefficients for Inference

The list of articles in the top three journals makes it clear that problems presenting MNL results are not limited to studies in international relations. For example, having found interaction effects of racially diverse contexts and racial stereotypes upon public policy, Weber, Lavine, Huddy and Federico (2014) examine the interaction of self-monitoring and racial context upon respondents' endorsements of stereotypes of African-Americans in order to better understand "why policy attitudes may be disconnected from racial stereotypes among high self-monitors" (71). This analysis provides another example of how particular baselines can create problems for interpreting MNL coefficients, particularly when using interaction terms.

Weber et al. (2014) consider that high self-monitors may be reluctant to endorse racial stereotypes in more racially-diverse contexts and use MNL to test explanations that high self-monitors in diverse contexts may be more likely to either opt-out from responding to questions, reject the stereotype, or choose the midpoint as a noncommittal response against a baseline of endorsing the stereotype. Tables 3 and 4 in Weber et al. (2014) present the MNL results for responses concerning stereotypes of blacks, respectively, as lazy and violent.

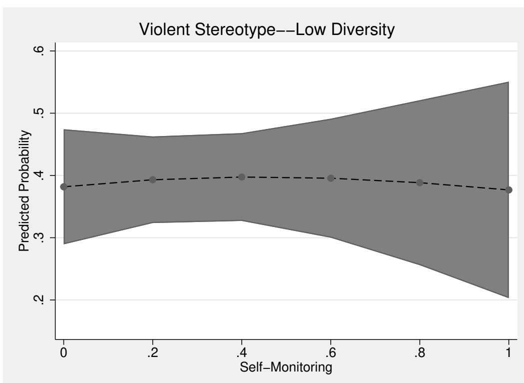
Weber et al. (2014) find clear evidence that high self-monitors are more likely to reject violent stereotypes in racially-diverse contexts, but they also argue that the interaction of self-monitoring and racial context influences choosing the midpoint response as a way of avoiding the question. For the lazy stereotype, they write that high self-monitors in highly diverse contexts are more likely to choose the midpoint; although, they refer to the estimated coefficient and note that this effect is not statistically significant (72). Examination of the predicted probabilities, however, indicates that self-monitoring had no effect upon choosing the midpoint. In high diversity contexts, low self-monitors had a predicted probability of .43 (se=.06) of choosing the midpoint, compared with high self-monitors' predicted probability of .42 (se=.10).

The paper then refers to results in Table 4 and discusses the effect of the interaction of racial context and self-monitoring upon choosing the midpoint for the violent stereotype: "self-monitoring also increases the likelihood of choosing the middle value of the scale, but only when diversity is high" (73). There is some ambiguity in this interpretation, but it does seem that the significant MNL coefficient motivates their conclusion. While they do not refer to the coefficient in making this statement, they also do not include, as they did with the lazy stereotype, the qualification that the effect is not statistically significant. Instead, the paper directs the reader to Figure 3, which shows the predicted increase (albeit without the confidence interval) in choosing the midpoint as self-monitoring increases only occurs in diverse contexts.

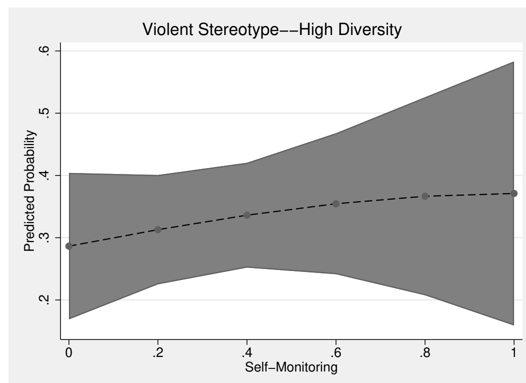
Figure A.1 reproduces only the effect of self-monitoring upon choosing the midpoint in different contexts and shows the increased probability of choosing the midpoint as self-monitoring increases for respondents in diverse contexts, but the difference in the predicted probability between low and high self-monitors is not statistically significant. The increase in the predicted probability of choosing the midpoint between the minimum and maximum value of self-monitoring is only .085, with an estimated standard error of .144 and 95% confidence interval of [-.199,.368]. The significant coefficient on the interaction relative to endorsing the stereotype is largely the result of the decrease in the predicted probability of endorsing the stereotype among high self-monitors in diverse contexts. Given this, it is difficult to conclude that this differences in the effect of self-monitoring in more diverse contexts choosing the midpoint can explain "why the connections between racial stereotypes of policy preferences are weak among these individuals" (73). Rather, Weber et al. (2014) are on much more solid ground in attributing the weaker connection to policy to high self-monitors rejecting the stereotype in high diversity contexts. It is also apparent from Figure A.1 that for any given level of self-monitoring, diversity does not have much effect upon choosing the midpoint. The maximum difference in predicted probabilities is for the lowest self-monitors, but that difference is only .095 (se=.071).

Figure A.1: Self-Monitoring, Diversity, and Violent Stereotype Midpoint Response

(a) Low Diversity



(b) High Diversity



Weber et al. (2014) present predicted probabilities as part of their analysis, but also use the coefficients for interpretation. Had Weber et al. (2014) used a different baseline, they would have found that the coefficient on the interaction was not significant. Perhaps the choice of "endorse" as the baseline was intended to determine whether high self-monitors in racially-diverse contexts who would otherwise choose to endorse the stereotype instead chose a more socially-acceptable response. The authors, however, did not focus upon the significant interaction coefficient for "opt-out," presumably because, as their Figure 3 makes very clear, the predicted probability of that option *drops* to .086 among high self-monitors from .130 for low self-monitors, a difference of .044 (se=.072). As with the Greenhill and Oppenheim (2017) examples, the sign and significance of the coefficient for one option can sometimes be a function of the change in the baseline probabilities relative to the corresponding change in some *other* option.

As other examples have illustrated, basing findings upon MNL coefficients can lead to misleading conclusions, and these problems are compounded with the well-known problems that arise when drawing inferences from the coefficients on interaction terms (Brambor, Clark and Golder 2006). Focusing upon MNL coefficients, as in the Weber et al. (2014) paper, can create ambiguity regarding the statistical basis of one's argument. As with all of the other examples in this paper, the predicted effects and their standard errors from MNL results provide a stronger basis for inference and a clearer way of communicating one's findings.

References

Brambor, Thomas, William Roberts Clark and Matt Golder. 2006. "Understanding interaction models: Improving empirical analyses." *Political analysis* 14(1):63–82.

Gelpi, Christopher. 2017. "The Surprising Robustness of Surprising Events: A Response to a Critique of Performing on Cue." *Journal of Conflict Resolution* 61(8):1816–1834. `https://u.osu.edu/gelpi.10/files/2016/08/GelpiPOCReplication-2clvxdn.zip`.

Greenhill, Kelly M. and Ben Oppenheim. 2017. "Rumor Has It: The Adoption of Unverified Information in Conflict Zones." *International Studies Quarterly* 61(3):660–676. `https://doi.org/10.7910/DVN/KMRVWY`, Harvard Dataverse, V1.

Weber, Christopher R., Howard Lavine, Leonie Huddy and Christopher M. Federico. 2014. "Placing Racial Stereotypes in Context: Social Desirability and the Politics of Racial Hostility." *American Journal of Political Science* 58(1):63–78. `https://doi.org/10.7910/DVN/DMIDY8`.