Adaptive Personality Inventories for Measuring Need for Cognition and Need to Evaluate

1. Introduction

Recent scholarship in political science has increasingly emphasized the role of personality traits in explaining public opinion and political behavior. This research has greatly expanded our understanding of how personality fundamentally structures decisionmaking, attitudes, and learning. However, political science literature has only scratched the surface in understanding the ways in which personality traits shape attitudes and behavior and how traits are in turn affected by political events. One explanation for this failure is surely that most established personality inventories contain far too many questions for inclusion on surveys. Standard practices in social and cognitive psychology result in evaluative batteries containing dozens or even hundreds of question items. For example, Cacioppo and Petty (1982) originally proposed a 40-item battery to measure need for cognition. Later, they proposed an "efficient" battery containing 18 questions chosen from the original 40 Cacioppo and Petty (1984).

Survey researchers typically avoid large scales. In part, this reflects the great financial burden of long surveys. Further, lengthy surveys drive up survey attrition, item-nonresponse, and even unit non-response. Long instruments also increase the burden on respondents who compensate by satisficing, increasing their use of "don't know" options, and providing less informative responses.[1]

The standard solution is to select a subset of available items, which are then administered to all respondents. These reduced scales are usually developed in one of three ways. First, scholars may examine the properties of the scale to make theoretically motivated decisions about which items to include. Another approach is to use the factor loadings in the original publication to select items. For instance, in designing a two-item battery measuring need for cognition for the American National Election Study (ANES), Bizer et al. (2004, p 13) chose "the two items that loaded most strongly on the latent construct in Cacioppoand Petty's (1982) factor analysis." (A nearly identical approach was taken in developing the three-item need to evaluate scale.) In this way, the need for cognition scale on the ANES is, in essence, based on the responses of 96 individuals drawn from

[1]Research on the effect of survey length on response quality includes: Heberlein and Baumgartner (1978); Herzog and Bachman (1981); Anderson, Basilevsky and Hum (1983); Burchell and Marsh (1992); Krosnick (1999); Crawford, Couper and Lamias (2001); Krosnick et al. (2002); and Galesic and Bosnjak (2009).

the faculty at the University of Iowa or workers on assembly lines in the Iowa City-Cedar Rapids area in the early 1980s (Cacioppo and Petty 1982, p 118-119). A final approach is to administer the battery to one or more fresh samples and to then use these new responses to choose a subset of existing items. For example, Muncer and Ling (2006) developed a 15-item reduced-form variant of the 40-item Empathy Quotient (Baron-Cohen et al. 2003) by analyzing responses from 362 students and parents at universities in North England.

In this document, I propose including a more efficient approach to handling large personality inventories on the ANES Pilot Study – Adaptive Personality Inventories (APIs). Specifically, I advocate including four-item APIs measuring need for cognition and need to evaluate. Below, I provide a brief overview of adaptive survey methods and provide experimental evidence supporting their utility in a survey setting. I conclude by providing specific implementation details.

2. Adaptive Personality Inventories

Generally, scholars developing reduced scales rely on estimates from calibration samples of some kind. Once a reduced inventory is chosen, it is administered to all respondents. APIs use computerized adaptive testing (CAT) algorithms to both reduce the amount of questions needed to measure the personality trait and increase the precision of measurement (Montgomery and Cutler 2013). APIs also rely on calibration samples to choose amongst potential survey items. However, APIs differ in that the goal is not to use this prior information to choose a single battery for all respondents, but rather to tailor the reduced battery to each respondent in a manner designed to maximize measurement precision.

2.1. **Algorithm essentials.**

Here, I briefly provide the details of *one* implementation that I apply below. APIs take a large population of potential items and select among them to efficiently place respondents on some latent scale. Roughly speaking, the algorithm chooses items that match the respondent's position on the trait (i.e., the respondent is likely to answer in one of several response categories) and items that are highly discriminatory.

The basic elements of an adaptive personality inventory are shown in Table 1 (Segall 2005, p 4). First, estimates ($\hat{\theta}_j$) are generated for each respondent's position on the trait (θ_j). Before the first item is administered, this estimate is based on our prior assumptions about θ_j. (We assume a common prior for all respondents, $\theta_j \sim \pi(\theta)$.) After each item in the inventory is administered, these expectations will be calculated based on both the prior and the respondent's answers to previously administered items.

Second, the next question item is selected out of the available battery. APIs choose the item that optimizes some pre-specified objective function. The third stage of the algorithm is to administer the chosen item and record the response. Fourth, the algorithm checks some stopping rule. In this case, the stopping rule is that the number of items asked of the respondent has reached some maximum value. Once the stopping criteria has been met, the algorithm produces final estimates of $\hat{\theta}_j$ and terminates.

TABLE 1. Basic elements of adaptive personality inventories

	Purpose	Description
1	Estimate positions	A provisional trait estimate, $\hat{\theta}_j$, is created based on first i responses. Alternatively, the estimate may be based on prior information.
2	Item selection	The item that optimizes some objective function is chosen.
3	Administer item	
4	Check stopping rule	Pre-defined stopping rules may include reducing posterior variance, $Var(\hat{\theta}_j)$, below a certain threshold or reaching some maximum time allotment for the battery.
5a	Repeat steps 1-4	If the stopping rule has not been reached, administer new items.
5b	Return estimate	If the stopping rule has been reached, calculate a final $\hat{\theta}_j$.

2.2. **Example implementation.** Although, there are a large number of variants of this basic algorithm in the literature, it is helpful to consider the details of one specific implementation.[2]

Most personality inventories include multiple response options. In an item response framework, this is often modeled using a graded response model (GRM). For each item i we assume that there are C_i response options. There is therefore a vector of threshold parameters defined as $\kappa_i = (\kappa_{i,0}, \kappa_{i,1}, \ldots, \kappa_{i,C_i})$, with $\kappa_{i,0} < \kappa_{i,1} \leq \kappa_{i,2} \leq \ldots, < \kappa_{i,C_i}$, $\kappa_{i,0} = -\infty$, and $\kappa_{i,C} = \infty$. In addition, each item is associated with a *discrimination* parameter a_i, which indicates how well each item corresponds to the underlying trait in question.

[2]Here I use the minimum expected posterior variance item selection criteria. In a review of competing item selection rules for the graded response model, Choi and Swartz (2009) note that this approach performs "equally well" to the more commonly used methods such as maximum posterior weighted information (MPWI), but that "the MEPV method would be preferred from a Bayesian perspective" (p. 18). Thus, my use of this method is more a reflection of taste than an indication that MEPV is in some way superior. Indeed, despite the large number of potential estimation methods, Choi and Swartz (2009) note that "for item banks with a small number of polytomous items, any of the methods ... are appropriate" (p. 18), which suggests that this choice is not critical in a survey setting.

To calculate the likelihood function, we need to estimate $P_{ijc} \equiv P(y_{ij} = c|\theta_j)$, which is the probability of answering in the c^{th} category for item i given the ability parameter for respondent j, denoted θ_j. This quantity, however, cannot be calculated directly. Instead, we need to define P^*_{ijc}, which is $\sum_{k=c+1}^{C_i} P_{ijk}$. Given these quantities, we calculate $P_{ijk} = P^*_{ij,k-1} - P^*_{ijk}$. Note that $P^*_{ij0} = 1$ and $P^*_{ijC_i} = 0$ in all cases. Under a logistic response assumption, therefore,

$$P^*_{ijk}(\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - \kappa_{ik})}}. \tag{1}$$

The likelihood function is, $L(\theta_j) = \prod_{i=1}^{n} \prod_{k=1}^{C_i} P_{ijk}^{I(y_{ij}=k)}$, where $I(.)$ is the indicator function.

2.2.1. *Calculating skill parameter.* Assuming that the respondent has answered $m-1$ items, calculating the expected a posteriori (EAP) estimate for the position of individual j given responses and the prior distribution $\pi(\theta_j)$ is,

$$\hat{\theta}_j^{(EAP)} = E(\theta_j|\mathbf{y}_{m-1,j}) = \frac{\int \theta_j \pi(\theta_j) L(\theta_j) d\theta_j}{\int \pi(\theta_j) L(\theta_j) d\theta_j}. \tag{2}$$

The posterior variance is then

$$Var(\theta_j) = E((\theta_j - \hat{\theta}_j^{(EAP)})^2|\mathbf{y}_{m-1,j})) = \frac{\int (\theta_j - \hat{\theta}_j^{(EAP)})^2 \pi(\theta_j) L(\theta_j) d\theta_j}{\int \pi(\theta_j) L(\theta_j) d\theta_j}. \tag{3}$$

As this involves a single dimension, we can estimate both using numerical integration.[3]

2.2.2. *Item selection.* In the example below, I use the minimum expected *a posteriori* (MEPV) item-selection criteria. To choose an item under this scheme, we follow three steps. First, we need to use the current estimate of $\hat{\theta}_j$ to estimate P_{ijk} for *each* possible response to *each* eligible (unasked) item using Equation 1. Second, we need to estimate $\hat{\theta}^{(EAP)*} = E(\theta_j|y_{k,m-1,j}, y^*_{mj})$, which is the EAP for individual j for all possible responses to all remaining question items. With this, we calculate $E((\theta_j - \hat{\theta}_j^{(EAP)*})^2|\mathbf{y}_{m-1,j}, y^*_{mj}))$, which is the posterior variance associated with each possible response to each remaining item. Third, we need to use these two elements to estimate the expected posterior variance for each item, which is $\sum_k P_{ijk} Var(\theta_j| \ldots, y^*_{ij} = k)$. This is the posterior variance we would have for each possible response to item i, weighted by the

[3]The software I have developed allows three different numerical integration routines so users can determine their optimal tradeoff in terms of speed and accuracy.

TABLE 2. Exemplar full and reduced-form measures of personality traits

	Original length	Reduced length
1998 ANES Pilot & ANES Time Series 2000-present		
Need for cognition	Cacioppo and Petty (1982)	Bizer et al. (2000)
	40	2
Need to evaluate	Jarvis and Petty (1996)	Bizer et al. (2000)
	16	3
American National Election Studies 2013 Internet followup		
Right wing authoritarianism	Altemeyer (1988)	
	30	5
Social dominance	Pratto et al. (1994)	
	15	2
Need for affect	Maio and Esses (2001)	
	26	4

probability of observing that response conditioned on our current estimate $\hat{\theta}_j$. We then select the item that minimizes this quantity.

2.2.3. *Stopping rule.* In this implementation, the algorithm stops offering items when the number of questions reaches a pre-specified threshold n_{max}. An alternative, however, is to stop when the posterior precision, $1/Var(\theta_j|\mathbf{y}_j)$, rises above some pre-specified level τ_θ^{stop}.

2.2.4. *Prior selection.* The prior distribution for θ_j is denoted $\pi(\theta_j)$. When using an "uninformative" prior, a natural choice is a conjugate normal prior $\pi(\theta_j) \sim N(\mu_\theta, \frac{1}{\tau_\theta})$, where τ_θ denotes the precision of the distribution. In the applications below we simply set $\tau = 1$, although any reasonably diffuse prior provides similar results.

3. EMPIRICAL EXPERIMENT

In this section, I summarize an experiment conducted using convenience samples recruited via Amazon's Mechanical Turk service. In the fall of 2014, I administered full-length versions of five personality inventories that have been included in reduced forms on the ANES to 1,204 subjects. The batteries were need for cognition (NFC), need to evaluate (NTE), need for affect (NFA), social dominance orientation (SDO)[4], and right wing authoritarianism (RWA) (see Table 2). Using these responses, I calibrated an API for each inventory as described above.

In the spring of 2015, I then recruited 1,335 new respondents who were randomly assigned to receive either fixed-reduced battery as used by the ANES or an API of the same length.[5] After

[4]I used only items measuring dominance attitudes (Peña and Sidanius 2002).
[5]Random assignment occurred once before each battery was administered.

completing the reduced battery, all subjects then answered all remaining questions in the full battery in a random order. I estimate respondents' scores using the full battery and treat them as respondents' "true" positions on the latent scale. I then compare the scores calculated using only questions selected by the fixed batteries and the API using these "true" positions as a common benchmark.

Table 3 shows the root mean squared error (RMSE)[6] for respondents answering either the fixed or adaptive scales.[7] As can be seen, the APIs provided more accuracy than the fixed batteries except in the case of NFC. (The NFC battery on the ANES is only two items long, and is therefore not well suited for APIs.) These results show that APIs provide more accurate estimates than widely used fixed batteries, even when there is only space for a few items.

TABLE 3. RMSE for adaptive vs. fixed

	Inventory name				
	NFA	NTE	NFC	SDO	RWA
Adaptive	0.47	0.47	0.49	0.36	0.44
Fixed	0.56	0.55	0.49	0.40	0.48

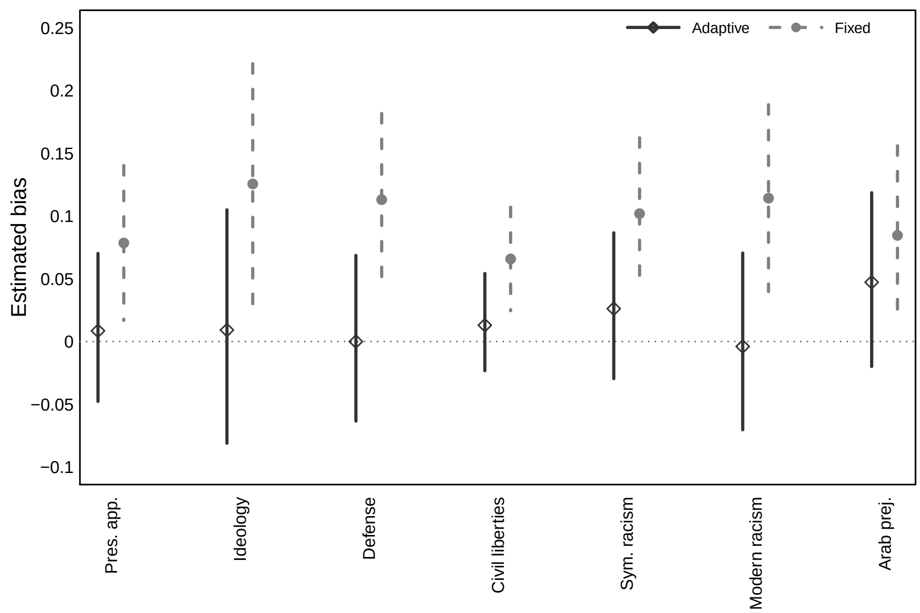
N=1,335

This improved accuracy has important consequences beyond mere measurement. To show this, I focus on the RWA measure, which originally had 30 items but was reduced to five on the ANES 2013 Internet followup study. Preliminary evaluations of the results showed that by failing to accurately recover the position of more evaluatively extreme individuals, the *fixed* RWA battery is left-censoring the estimates. This, in turn, biases our understanding for how RWA relates to other important factors. To illustrate this, I included measures of several constructs theoretically related to RWA, including presidential approval, ideology, defense spending attitudes, civil liberties attitudes, symbolic racism, modern racism, and prejudice towards Arabs and Muslims. I estimated separate regressions by treatment condition using RWA as an explanatory variable and these other constructs as dependent variables.[8] I then estimated the "true" value for these regression coefficients using respondents' scores as estimated from the full battery. I calculate the bias as the difference between the regression coefficient estimated using the reduced and full batteries. The

[6]Recall that we use estimates generated from using the entire battery as their "true" position (θ_j). For observation j, squared error is $((\hat{\theta}_j - \theta_j)^2)$.
[7]Estimates were generated using a GRM model fit only with the second sample.
[8]In these regressions, I control only for race, gender, and level of education.

results, shown in Figure 1, illustrate clearly that the censored measure of RWA from the fixed battery upwardly biases these regression coefficients, leading us to conclude that RWA is a stronger predictor of these constructs than is actually the case. However, this bias is less prevalent when examining the reduced *adaptive* RWA scale.

FIGURE 1. Bias in regression estimates for RWA and seven related constructs



4. PROPOSAL: APIs FOR NEED FOR COGNITION AND NEED TO EVALUATE

I propose that the ANES Pilot Study include four-item adaptive versions of the need for cognition (NFC) and need to evaluate (NTE) scales as potential replacements for the fixed-reduced scales first introduced in the 1998 Pilot Study and included on the time series studies since 2000. Since their inclusion, these scales have been widely used in prominent articles in both political science and public opinion that focus on candidate choice (e.g., Druckman 2004; Bizer et al. 2004; Nir 2011; Rudolph 2011; Sokhey and McClurg 2012; Chong and Druckman 2013). The research shown above suggests that adaptive versions of these batteries should improve scholarly studies exploring the role of these important personaltiy traits play in affecting voters' decisions.

I will calibrate these APIs using a combination of convenience samples[9] and data from The American Panel Survey (TAPS). The latter is an ongoing nationally representative survey that has included the *full* NTE scale twice (Dec. 2014 and Feb. 2015) and the *full* 40-item NFC battery three times (Dec. 2014, Feb. 2015, and May 2015). Combining these data sources will allow me

[9]I have now administered these batteries to more than 4,000 individuals spanning three years as part of my research on adaptive personality inventories.

to ensure accurate item-parameter estimations while also allowing me to "norm" the battery to the TAPS sample so that our prior expectations are reasonable for the ANES sample.

With support from independent programming contractors, I have developed a webservice,[10] which is hosted on the Heroku cloud application platform. The webservice is designed to respond to queries from external servers and to execute the specified item selection routines. So, for instance, a survey run on the Qualtrics platform can call on the webservice to determine which question should be asked next given a specific response history.[11] Testing thus far in Qualtrics shows that incorporating the CAT algorithm in this manner results in virtually no delays with batteries as large as 60 items.[12] Further, the Heroku platform is fully scalable, allowing me to set up multiple instances such that speed is maintained even when many surveys occur simultaneously.[13]

4.1. **Need for Cognition.** For the following statements, please indicate your level of agreement or disagreement with each statement. *Response options: Strongly Agree, Agree, Uncertain, Disagree, Strongly Disagree*. (* indicates questions previously on the ANES.)

(1) I would prefer complex to simple problems.*
(2) I like to have the responsibility of handling a situation that requires a lot of thinking.*
(3) Thinking is not my idea of fun.
(4) I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.
(5) I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.
(6) I find satisfaction in deliberating hard and for long hours.
(7) I only think as hard as I have to.
(8) I prefer to think about small, daily projects to long-term ones.
(9) I like tasks that require little thought once Ive learned them.
(10) The idea of relying on thought to make my way to the top appeals to me.
(11) I really enjoy a task that involves coming up with new solutions to problems.
(12) Learning new ways to think doesnt excite me very much.
(13) I prefer my life to be filled with puzzles that I must solve.
(14) The notion of thinking abstractly is appealing to me.
(15) I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.

[10]The webservice is based on the `catSurv` R package, which implements the various item-selection routines described in the literature. This software wass designed from the ground up to support fast real-time computation for large scale survey projects, and item selection routines are implemented in C++.
[11]The CAT model – including options for ability estimation, priors, item selection routines, and the response history – is passed in as a JSON object. The server then calculates which questions should be administered next.
[12]This is facilitated by pre-calculating next steps while the survey taker reads and answers the current question.
[13]An alternative solution, which I implemented in collaborationg with Gfk/Knowledge networks, is to pre-calculate the possible branchings based on all potential responses to all questions. For short batteries (3-5 items), this is also easily included in a Qualtrics survey.

(16) I feel relief rather than satisfaction after completing a task that required a lot of mental effort.
(17) Its enough for me that something gets the job done; I dont care how or why it works.
(18) I usually end up deliberating about issues even when they do not affect me personally.

4.2. **Need to Evaluate.** How well does each of the following items describe you? *Response options: Extremely Characteristic, Somewhat Characteristic, Uncertain, Somewhat Uncharacteristic, and Extremely Uncharacteristic.* (* indicates questions previously on the ANES.)

(1) It is very important to me to hold strong opinions.
(2) I like to have strong opinions even when I am not personally involved.
(3) I would rather have a strong opinion than no opinion at all.
(4) I form opinions about everything.*
(5) I have many more opinions than the average person.*
(6) I enjoy strongly liking and disliking new things.
(7) I often prefer to remain neutral about complex issues.*
(8) I only form strong opinions when I have to.
(9) It bothers me to remain neutral.
(10) I pay a lot of attention to whether things are good or bad.
(11) I want to know exactly what is good and bad about everything.
(12) I am pretty much indifferent to many important issues.
(13) I prefer to avoid taking extreme positions.
(14) There are many things for which I do not have a preference.
(15) I like to decide that things are really good or really bad.
(16) If something does not affect me, I do not usually determine if it is good or bad.

References

Altemeyer, Bob. 1988. *Enemies of Freedom: Understanding Right-Wing Authoritarianism*. San Francisco, CA: Jossey-Bass.

Anderson, A.B., A. Basilevsky and D. Hum. 1983. Missing Data: A Review of the Literature. In *Handboook of Survey Research*, ed. Peter H. Rossi, James D. Wright and Andy B. Anderson. New York: Academic Press pp. 415–481.

Baron-Cohen, Simon, Jennifer Richler, Dheraj Bisarya, Nhishanth Gurunathan and Sally Wheelwright. 2003. "The systemizing quotient: an investigation of adults with Asperger syndrome or high–functioning autism, and normal sex differences." *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358(1430):361–374.

Bizer, George Y, Jon A Krosnick, Allyson L Holbrook, S Christian Wheeler, Derek D Rucker and Richard E Petty. 2004. "The impact of personality on cognitive, behavioral, and affective political processes: The effects of need to evaluate." *Journal of Personality* 72(5):995–1028.

Bizer, George Y, Jon A Krosnick, Richard E Petty, Derek D Rucker and S Christian Wheeler. 2000. "Need for cognition and need to evaluate in the 1998 National Election Survey Pilot Study." National Election Studies Report.

Burchell, Brendan and Catherine Marsh. 1992. "The effect of questionnaire length on survey response." *Quality & Quantity* 26(3):233–244.

Cacioppo, John T. and Richard E. Petty. 1982. "The Need for Cognition." *Journal of Personality & Social Psychology* 42(1):116–131.

Cacioppo, John T. and Richard E. Petty. 1984. "The Efficient Assessment of Need for Cognition." *Journal of Personality Assessment* 48(3):306–307.

Choi, Seung W. and Richard J. Swartz. 2009. "Comparison of CAT item selection criteria for polytomous items." *Applied Psychological Measurement* 33(6):419–440.

Chong, Dennis and James N Druckman. 2013. "Counterframing effects." *Journal of Politics* 75(01):1–16.

Crawford, Scott D., Mick P. Couper and Mark J. Lamias. 2001. "Web Surveys : Perceptions of burden." *Social Science Computer Review* 19(2):146–162.

Druckman, James N. 2004. "Political preference formation: Competition, deliberation, and the (ir)relevance of framing effects." *American Political Science Review* 98(4):671–686.

Galesic, Mirta and Michael Bosnjak. 2009. "Effects of questionnaire length on participation and indicators of response quality in web surveys." *Public Opinion Quarterly* 73(2):349–360.

Heberlein, Thomas A. and Robert Baumgartner. 1978. "Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature." *American Sociological Review* 43(4):447–462.

Herzog, A. Regula and Jerald G. Bachman. 1981. "Effects of questionnaire length on response quality." *Public Opinion Quarterly* 45(4):549–559.

Jarvis, W Blair G and Richard E Petty. 1996. "The need to evaluate." *Journal of Personality and Social Psychology* 70(1):172–194.

Krosnick, Jon A. 1999. "Survey Research." *Annual Review of Psychology* 50:537–67.

Krosnick, Jon A., Allyson L. Holbrook, Matthew .K. Berent, Richard A. Brody T. Carson, W.Michael Hanemann, Raymond J. Kopp, C. Mitchell, Robert Cameron, Stanley Presser, Paul A. Ruud, V.Kerry Smith, Wendy R. Moody, Melanie C. Green and Michael Conaway. 2002. "The impact of 'no opinion' response options on data quality: Non-attitude reduction or an invitation to satisfice?" *Public Opinion Quarterly* 66(3):371–403.

Maio, Gregory R and Victoria M Esses. 2001. "The need for affect: Individual differences in the motivation to approach or avoid emotions." *Journal of Personality* 69(4):583–614.

Montgomery, Jacob M. and Josh Cutler. 2013. "Computerized adaptive testing for public opinion surveys." *Political Analysis* 21(2):141–171.

Muncer, Steven J and Jonathan Ling. 2006. "Psychometric analysis of the empathy quotient (EQ) scale." *Personality and Individual Differences* 40(6):1111–1119.

Nir, Lilach. 2011. "Motivated reasoning and public opinion perception." *Public Opinion Quarterly* 75(3):504–532.

Peña, Yesilernis and Jim Sidanius. 2002. "US patriotism and ideologies of group dominance: A tale of asymmetry." *The Journal of social psychology* 142(6):782–790.

Pratto, F., J. Sidanius, L.M. Stallworth and B.F. Malle. 1994. "Social dominance orientation: A personality variable predicting social and political attitudes." *Journal of Personality and Social Psychology* 67(4):741–741.

Rudolph, Thomas J. 2011. "The dynamics of ambivalence." *American Journal of Political Science* 55(3):561–573.

Segall, Daniel O. 2005. Computerized Adaptive Testing. In *Encyclopedia of Social Measurement*. Vol. 1 Oxford: Elsevier pp. 429–438.

Sokhey, Anand Edward and Scott D McClurg. 2012. "Social networks and correct voting." *The Journal of Politics* 74(03):751–764.