Selling ourselves short? How abbreviated measures of personality change the way we think about personality and politics

Bert N. Bakker and Yphtach Lelkes[1]

University of Amsterdam & University of Pennsylvania

Accepted for publication in The Journal of Politics

[1]Supplementary material for this article is available in the appendix in the online edition. Replication files are available in the JOP Data Archive on Dataverse (https://dataverse.harvard.edu/dataverse/jop). Study 1 was reviewed and approved by the Institutional Review Board at the University of Amsterdam. The data for Study 2 and 3 were collected by CentERdata (Tilburg University). CentERdata abides by the Dutch "protection of personal data act", which is consistent with and derived from European law (Directive 95/46/EC). Support for this research (Study 1) was provided by the Amsterdam School of Communication Research, University of Amsterdam.

Abstract

Political scientists who study the interplay between personality and politics overwhelmingly rely on short personality scales. We explore whether the length of the employed personality scales affects the criterion validity of the scales. We show that Need for Cognition (NfC) increases reliance on party cues, but only when a longer measure is employed. Additionally, while NfC increases reliance on policy information, the effect is more than twice as large when a longer measure is used. Finally, Big Five personality traits that have been dismissed as irrelevant to political ideology yield stronger and more consistent associations when larger batteries are employed. We also show that using high Cronbach's alpha and factor loadings as indicators of scale quality does not improve the criterion validity of brief measures. Hence, the measurement of personality conditions the conclusions we draw about the role of personality in politics.

Keywords: Personality, Measurement, Need for Cognition, Big Five

Selling ourselves short? How abbreviated measures of personality change the way we think about personality and politics

The study of personality and politics has – after lying relatively dormant for several decades – received renewed interest from political scientists (e.g., Bakker, Rooduijn, & Schumacher, 2016; Bullock, 2011; Feldman & Johnston, 2014; Gerber, Huber, Doherty, Dowling, & Ha, 2010; Malka, Soto, Inzlicht, & Lelkes, 2014; Mondak & Halperin, 2008). We therefore see more and more personality measures being included in political science research. Yet, space is scarce in omnibus surveys – such as the American National Election Studies or the British National Election Study – that are the source for much of the personality and politics research. Furthermore, researchers often have limited resources at hand when designing their own studies. The trade-off for limited space and resources are shorter measures. While textbooks on measurement recommend long scales over short scales (Cronbach, 1949), political science research tends to ignore this advice (although see, Achen, 1975; Ansolabehere, Rodden, & Snyder, 2008).

We turn our attention to personality traits, which are often measured using one or two items culled from fairly long and often multi-dimensional scales. As is well known, such short forms are less reliable (Gosling, Rentfrow, & Swann, 2003; Schmitt, 1996) and may measure only some sub-dimension of a trait (Cronbach, 1949; Smith, McCarthy, & Anderson, 2000) leading to either regression dilution or overestimation of the association between a trait and a criterion measure (Credé, Harms, Niehorster, & Gaye-Valentine, 2012; Lord & Novick, 1968; Smith et al., 2000). Despite these risks, short personality measures continue to be a mainstay of personality and politics research in political science.

In this paper, we assess the impact of the short measures on substantive conclusions. We focus on two central debates within the literature. The first touches upon the role of Need for Cognition (NfC) – the tendency to enjoy thinking (Cacioppo & Petty, 1982) – in moderating the reliance upon party cues and policy information (Bullock, 2011; Kam, 2005). The second addresses the association between the Big Five

personality traits and political ideology (Gerber et al., 2010; Mondak & Halperin, 2008). In our studies, we show that the brief measures yield different results than the longer measures. We also find that a slight increase in the number of items we use yields outcomes consistent with the more elaborate measures, which also implies that political science research need not turn to excessively long measures of personality.

Our paper has important substantive implications. First, although we replicate the main findings of Kam (2005), unlike that paper, we find that NfC does moderate the reliance on party cues in a way that is consistent with theories of motivated reasoning (Kahan, 2012; Petersen, Skov, Serritzlew, & Ramsøy, 2013), i.e., those higher on NfC rely more upon party cues. This finding contradicts theories of bounded rationality (Popkin, 1994) which suggests that a party cue is an easy heuristic for those low on NfC. We only reach these conclusions if we rely upon the full 18-item battery, while abbreviated batteries – such as a 2-item NfC measure (developed by Bizer, Krosnick, Petty, Rucker, & Wheeler, 2000) – would lead us in line with Kam (2005) to conclude that NfC *does not* moderate the reliance upon party cues. Next, we turn to the reliance upon policy information. The majority of political science studies have relied upon an abbreviated 2-item NfC measure and do not find evidence that those higher on NfC rely *more* upon policy information (Berinsky & Kinder, 2006; Holbrook, 2006; Mérola & Hitt, 2016; Rudolph, 2011; Sokhey & McClurg, 2012). We show that this conclusion is an artifact of the measure – using a full 18-item battery, NfC in fact moderates the reliance upon policy information as would be predicted based the Eleboration Likelihood Model.

In a third study, we show that, if we rely upon a brief measure of the Big Five traits (i.e., 10-item Big Five Inventory; Rammstedt & John, 2007), then we conclude that traits such as Agreeableness, Extraversion and Conscientiousness are irrelevant for politics, while other traits are weakly associated with political attitudes. Yet, once we rely upon a more elaborate battery (i.e., the 50-item IPIP-FFM; Goldberg et al., 2006), many of the Big Five personality are as highly correlated with the same political outcomes as Openness, the trait perhaps most commonly shown to be important for

politics.

The consequences of using brief personality measures

Brief measures of personality offer several advantages over their longer counterparts. Short measures of personality are cheaper to administer, increase response rates (Edwards, Roberts, Sandercock, & Frost, 2004) and decrease measurement error that may arise due to boredom and fatigue caused by completing a long personality battery (Burisch, 1984). Since hundreds of questions often appear in a single wave of an omnibus survey, space comes at a premium, and brief measures allow scholars to study personality when they have limited space available in a survey. Fortunately, some of the psychometric properties of brief personality measures are satisfactory. For the NfC, Big Five Inventory (henceforth, BFI; Rammstedt & John, 2007) and other brief measures of personality the test-retest reliability (Gerber, Huber, Doherty, & Dowling, 2013; Gosling et al., 2003; Rammstedt & John, 2007) and the convergent validity – i.e., the degree to which the brief measure correlates with a longer measure of the same construct – are acceptable (Donnellan, Oswald, Baird, & Lucas, 2006; Rammstedt & John, 2007).

Other psychometric properties of brief personality batteries are more problematic. First, in all domains of research, it's well known that short batteries tend to be less reliable than longer batteries (Lord & Novick, 1968). The measurement error in the independent variables attenuates the relationship with certain criterion (i.e., outcome) measures – a process called regression dilution.

Second, personality research is particularly sensitive to short batteries because of the so-called "bandwidth-fidelity" trade-off (Cronbach & Gleser, 1965). Bandwidth is the amount of complexity of information in a measure. Many personality traits are fairly complex constructs and thus have a high bandwidth. Each of the Big Five traits (Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism) consists – according the Five Factor Model of personality – of six lower sub-dimensions. Conscientiousness, for instance, contains the sub-dimensions Achievement Striving, Competence, Dutifulness, Deliberation, Self-Discipline and Order (Costa, McCrae, &

Dye, 1991). Short batteries, such as the BFI and the Ten Item Personality Inventory (henceforth, TIPI; Gosling et al., 2003) may only tap into a few of these sub-dimensions. If the target measure we are interested in is associated with a sub-dimension that is missed by the BFI or TIPI, any correlation will be attenuated (Credé et al., 2012). Conversely, if a target measure is only related to one specific facet of that trait, but not others, any correlation will be overestimated (Credé et al., 2012). Accordingly, there is the risk of a Type M (magnitude) error (Gelman & Carlin, 2014). A Type S (sign) error might even occur if our brief measure disproportionally taps into a sub-dimension that is differentially correlated with the criterion measure than the broader trait.

Short measures with low bandwidth: The case of the Need for Cognition and message processing

NfC captures individual differences in the tendency to enjoy thinking (Cacioppo & Petty, 1982). It is possible that NfC may either increase *or* decrease the reliance on party cues compared to the same information without party cues. On the one hand, in line with the expectation by Kam (2005), party cues offer an easy heuristic to those who would prefer not to think about the implications about a policy. Therefore, we could expect that reliance on party cues is higher among those who are lower on NfC. However, those who are "the most vulnerable to ideologically consistent bias" (Hatemi & McDermott, 2016, p. 342) tend to be the cognitively reflective and politically sophisticated (Kahan, 2012; Slothuus & De Vreese, 2010), two traits that are positively correlated with the NfC (Kahan, 2012; Tidwell, Sadowski, & Pate, 2000). Therefore, among those higher on NfC party cues may also trigger motivated reasoning (Petersen et al., 2013) and lead people to conform to identity-consistent attitudes (Kahan, 2012; Slothuus & De Vreese, 2010).

Aside from party cues, the literature is fairly clear on the expectation that NfC should increase the reliance on substantive information. The Elaboration Likelihood Model (Petty & Cacioppo, 1986), which underpins much of the political persuasion literature (Alvarez & Brehm, 1995; Johnson & Martin, 1998), argues that individual

differences in NfC moderate the extent to which citizens' political attitudes are influenced by policy information. Those high on NfC tend be motivated to understand and thoroughly process information that they receive and, therefore, should be more affected by policy information compared to citizens that score low on NfC (Bullock, 2011).

When testing the effects of NfC on information processing, political scientists tend to rely on a two-item variant which was originally developed for inclusion in the 2000 American National Election Studies (henceforth, ANES; Bizer et al., 2000). The two item ANES NfC measure is a highly shortened form of the 18-item NfC scale (Cacioppo, Petty, & Kao, 1984), which itself is a "short" form of the original 34-item scale (Cacioppo & Petty, 1982). Studies within political science that rely upon the 2000 ANES measure often fail to find evidence that NfC moderates the impact of party cues (Bullock, 2011; Kam, 2005) or policy information (Berinsky & Kinder, 2006; Holbrook, 2006; Mérola & Hitt, 2016; Rudolph, 2011; Sokhey & McClurg, 2012) on political attitudes. Bullock (2011, p.513) suggests that "the accumulating non-findings about NfC may well be driven by measurement error." Employing a somewhat larger (although not validated) 6-item NfC battery Bullock (2011) finds that NfC does moderate the effect of policy information on policy attitudes. Yet NfC does not moderate the effect of policy information when Bullock (2011) subsetted the 6-item battery to the 2-item ANES battery. Importantly, Bullock (2011) does not compare the 6-item measure to a validated NfC battery, so we have to be cautious in over-interpreting these results.

We conduct two studies to determine whether the measurement of the NfC conditions the conclusions we reach about the extent to which the NfC moderates citizens' tendency to rely upon party cues (Study 1) and policy information (Study 2). We have formulated competing expectations of the nature through which the NfC moderate the reliance upon party cues, while we expect more reliance upon policy information by persons high on NfC. Yet in both studies we expect that the conclusion we draw about the importance of the NfC is conditional upon the measurement of the NfC.

Study 1: Need for Cognition and Party Cues

In study 1, we test whether the Need for Cognition conditions the reliance upon party cues (Bullock, 2011; Kam, 2005). Specifically, we can assess whether those high on the NfC rely more (Kahan, 2012) or less on party cues (Kam, 2005). In order to test these competing expectations, we replicate one of the most influential papers that assesses the extent to which the reliance upon party cues is moderated by the NfC conducted by Kam (2005).[2] Participants were randomly exposed to a short newspaper article introducing a proposal to ban food irradiation – a low salience political issue in the United States. In the first condition, Democrats supported the policy, and Republicans opposed it. In the second condition, Republicans supported the policy, while Democrats opposed it. In the control condition no party cues were mentioned, while all other information remained constant.[3] We implemented the design directly in line with Kam (2005).[4] Survey Sampling International (SSI) fielded the experiment in the United States on their on-line panel between July 4 and July 6, 2016. In exchange for participation, SSI rewards panelists with points that can be exchanged for various rewards. In total 883 respondents were randomly assigned to participate in the cue-taking experiment.

Following Kam (2005), we measured support for the ban on food irradiation on a five point Likert-scale scored on a scale ranging from "strongly agree" (1) to "strongly disagree" (5). In order to decrease measurement error in the dependent variable, we included two additional items, namely "The costs of food irradiation outweigh the benefits" which was scored ranging on a scale from "strongly agree" (1) through "strongly disagree"(5) and "All things considered, food irradiation is a good thing" scored on a scale ranging from "food irradiation is bad" (1) through "food irradiation is good"(5). We created a additive scale ranging from (0) oppose a ban on food irradiation

[2]The paper had received 287 citations by September 5, 2017 (Google Scholar).

[3]Appendix A.1 provides stimuli.

[4]Kam (2005) conducted her study in Michigan and focused upon the House of Representatives in the state of Michigan. We conducted our study across the United States and accordingly the proposed ban on food irradiation as is discussed in the House of Representatives.

to (1) support a ban on food irradiation (M=0.54, SD=0.20, α=0.54).

The NfC was measured with the 18-item battery (Cacioppo et al., 1984) using items such as "I prefer complex to simple problems" which respondents answer on a likert-type scale from "strongly disagree" (1) to "strongly agree (7)". We compute the average score of the measures (on the original 7 point scale), and then, for the sake of comparability, rescale to lie between 0 and 1 by subtracting the minimum score (1), and then dividing by the maximum minus the minimum (6). The 18-item NfC battery can be subsetted to the 2-item measure that is included in the ANES 2000 (Bizer et al., 2000) as well as the 6-item battery employed by Bullock (2011). These NfC batteries – as well as other personality inventories in this paper – were rescaled to range from 0 to 1 like we did with the 18-item battery. Appendix A.2 provides descriptive statistics and psychometric properties of the scales. Randomization checks indicated that NfC as well as a set of observed background characteristics were balanced across the treatments (Appendix A.3).

Treatment status was indicated by a set of dummy variables indicating if the participant read that the party they identify with proposed to ban food irradiation (In-Party Cue) or opposed the ban (Out-Party Cue). A control condition omitted mention of the party label (No Party Cue).

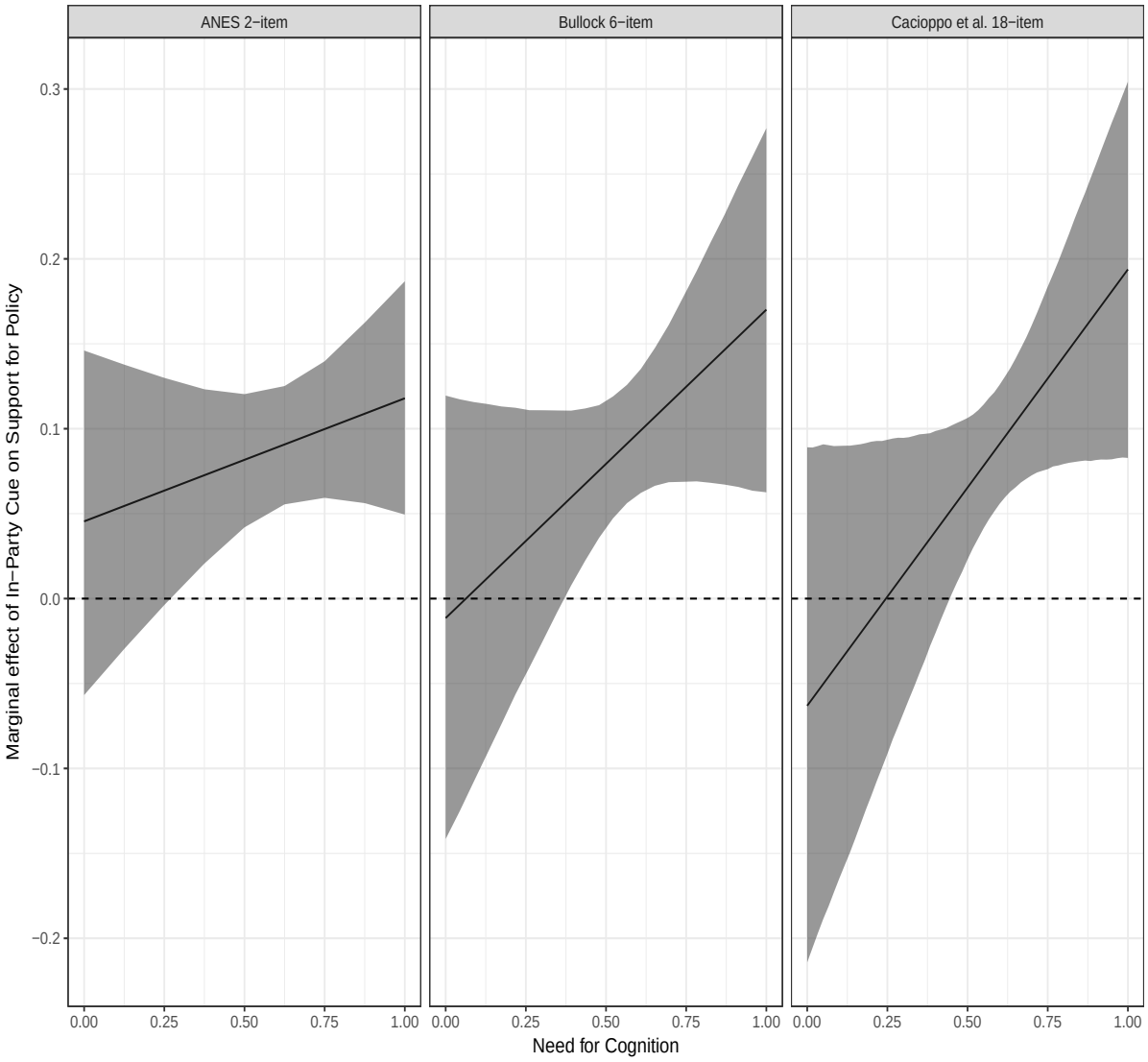
While this study replicates the main effect findings of Kam (2005) that citizens rely upon party cues (Appendix A.4), we are interested in the question whether the NfC moderates the reliance upon the party cues. To investigate when the reliance upon party cues is moderated by the NfC, we estimated three OLS regression models. We set the *Out-party Cue* as the reference category and interacted the NfC with the *In-Party Cue* and *No-party Cue*. Figure 1 presents the results of the regression models and plots the marginal effect of the In-Party Cue on support for banning food irradiation over the range of the NfC. Note that here and throughout the manuscript we report unstandardized coefficients; standardized coefficients appear in the appendix.

In line with Kam (2005), we find that using the two item ANES measure (left-hand column), the NfC *does not* moderate the effect of the party cue on the policy

attitude (b=0.08, SE=0.07). NfC does moderate the reliance upon party cues when we use an 18-item NfC measure (right-hand column; b=0.26*, SE=0.13). In line with motivated reasoning, we find that citizens that score high on the NfC rely more upon the In-Party Cue compared to the Out-party Cue (Kahan, 2012; Slothuus & De Vreese, 2010). The 6-item measure does not significantly moderate the reliance upon party cues (b=0.18, SE=0.11), however, the effect is more than twice as large when using the two-item measure. The results do not change once we run structural equation models (Appendix B.6), rely upon the single item dependent variable as employed by Kam (2005) (Appendix B.7), or employ the slightly different operationalization of the treatment indicators found in Appendix A.8 (Appendix B.8).

Our results might be due to the idiosyncrasies of the 2-item ANES measure. Perhaps another brief NfC inventory results in estimates more consistent with a larger NfC battery. To estimate whether the effect size is a function of the particular items or the number of items included in the battery, we turn to a second set of analysis. Using the 18 items, we generated all possible combinations for scales of different length. This resulted, for instance, in 153 2-item measures, 816 3-item measures, 3,060 4-item measures, 8,586 5-item measures, 18,564 6-item measures, etc. For each of these 262,143 instantiations of NfC, we then calculated the interaction effects between receiving the In-Party cue or Out-Party cue and NfC. Figure 2 plots the distribution of the point estimates of these measures sorted by the number of items used to generate the trait. The x-axis in each panel indicates the number of items used to make a particular trait. The median point estimate is plotted as the thick horizontal in each boxplot. The results show that a randomly generated two item scale yields estimates of an interaction effect that is roughly 2.5 times smaller compared to the 18 item scale. The median point estimate of randomly chosen scales does not increase linearly with the number of items in a scale, however. The median point estimate of a randomly-chosen four item scale is three-fourths the size of the 18-item scale, by 9 items, the median coefficient estimate that is roughly 90 percent of the 18-item NfC estimate. Study 1 shows that the ANES measure seems to yield estimates that are smaller than roughly 70 percent of

Figure 1. Study 1: Need for Cognition and party cues



Note: Marginal effects of the In-Party cue compared to the Out-Party cue on support for food irradiation with 95% Confidence Intervals are plotted (see Appendix A.5 for tables with results).

any other two-item measure.

Figure 2. Study 1: Relationship between interaction effect size and the number of items used to create NfC

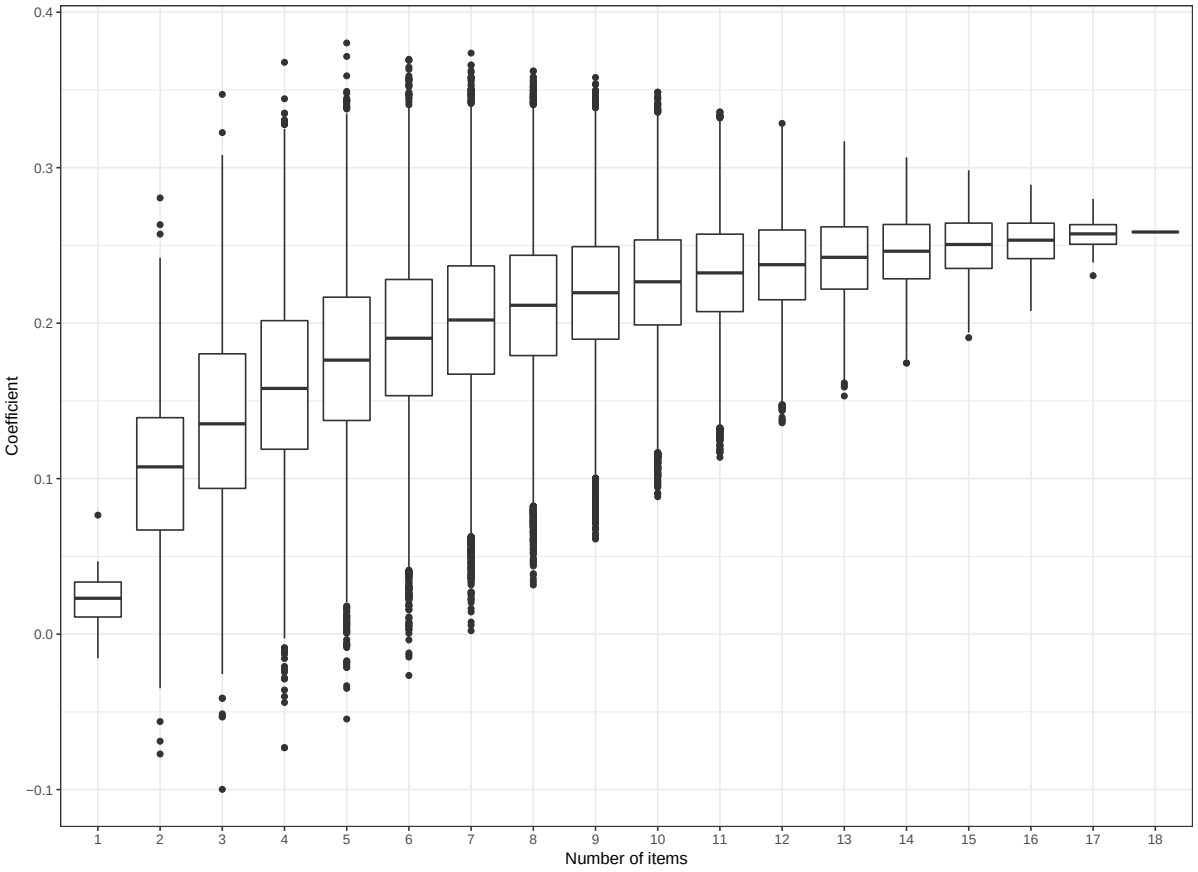


Figure plots the coefficient of the interaction effect between NfC and the policy cues and each possible combinations of the NfC. Distribution of the point estimates of these measures sorted by the number of items used to generate the NfC.

Could short measures be saved if we just more carefully construct them? For low bandwidth scales, one suggestion is to make sure that the Cronbach's alpha is at least "satisfactory" (greater than .7).[5] We show, that Cronbach's alpha increases with the number of items, as expected, but when we compare scales that contain the same number of items, there is no association between the Cronbach's alpha of a scale and the extent to which estimates are closer to the 18-item NfC measure (Appendix B.9).

[5]Bullock (2011, p.503), for instance, suggested that the low Cronbach's alpha of the 2-item ANES measure explain the poor performance of the short battery.

Another common practice in selecting items to form a short measure is to select items that load most highly in a confirmatory factor analysis. This was the method used to develop the 2-item NfC ANES measure (Bizer et al., 2000). Yet, we find no association between the extent to which items load high on a brief NfC measure and closeness to the estimate of the 18-item measure (Appendix A.10). This suggests that relying solely upon Cronbach's alpha or high factor item loadings does not result in brief measures of personality with a high criterion validity. Instead, Study 1 suggests that decreasing measurement error by increasing a number of items is the best strategy.

Study 2: Need for Cognition and Policy Information

In study 2, we test whether NfC moderates the effect of policy cues on policy attitudes. The survey experiment – designed by a separate team (Coffe, 2013) – was a 2 (Policy Cue: center-right vs. radical-right message) X 2 (Ideological cue: no ideological cue vs. ideological cue) X 2 (Sex: male vs. female politician) X 2 (Speech: aggressive vs. nuanced manner of speech) experimental design. Participants were shown a professionally edited campaign video in which they were randomly assigned to a center-right message or radical-right message. For instance, discussing the issue of immigration, the *center-right* message was "we believe that the influx of underprivileged, lowly educated immigrants must stop. Instead, we should open our doors only to higher educated, promising immigrants," while the *radical-right* message was, "We demand a complete cessation of immigration of people from Islamic countries."[6] Each treatment lasted approximately two minutes. Our interest lies in the extent to which participants rely upon policy information (center-right vs. radical-right message).

The experiment was run on the Longitudinal Internet Studies for the Social Sciences (LISS) panel, a true probability sample of Dutch households drawn from the official population registry (Scherpenzeel & Das, 2010) and administered by CentERdata (Tilburg University). LISS panel members fill out one survey each month and get reimbursed 15 Euro per hour. We used three waves of the LISS panel. Between

[6]Appendix B.1 provides complete treatments.

October 1 and October 30, 2012 6,434 LISS panel members were invited to participate in a survey which contained a cue-taking experiment and 5,179 panel members completed the survey (80.5% response rate). In total 455 (i.e., 8.77%) participants indicated that they could not hear and/or see the videos in the experiment. These participants were excluded from further analyses.[7]

The dependent variable measured the extent to which citizens agree with the party position, namely "How much do you agree with the party position on migration by the candidate of the political party?" which was scored on scale ranging from "completely disagree" (1) through "completely agree" (5). We recoded the scale to range from 0 (completely disagree) to 1 (completely agree).

NfC was measured as part of the "Personality and Values module" of the LISS panel which is fielded to LISS panel members in May of each year; to increase our sample size, we include NfC from May 2011 (74% response rate) and May 2012 (79.3% response rate). By merging these waves, we had a measure of the Need for Cognition for almost all participants (N=4,554). The 18-item NfC battery can be subsetted to the ANES 2-item measure (Bizer et al., 2000) as well as the 6-item battery employed by Bullock (2011). Appendix B.2 provides the item wording of the NfC, descriptive statistics and psychometric properties of the three NfC measures. Randomization checks indicate that the three NfC batteries are randomly distributed across the different treatments (Appendix B.3).

Figure 3 presents the results of the regression models and plots the marginal effect of the radical-right message on support for the proposed migration policy over the range of the NfC. We control for the other conditions in the experiment and the interaction between the NfC and these conditions. Using the two item ANES measure, we find that NfC moderates the effect of the policy cue on policy attitudes (b=-0.08, SE=0.04). That is, the agreement with the radical-right policy decreases compared to the right-wing policy as NfC goes up (see left-hand panel of Figure 3). However, when running a

[7]The tendency to not be able to hear and/or see the treatment is randomly distributed across the conditions (Appendix B.3).

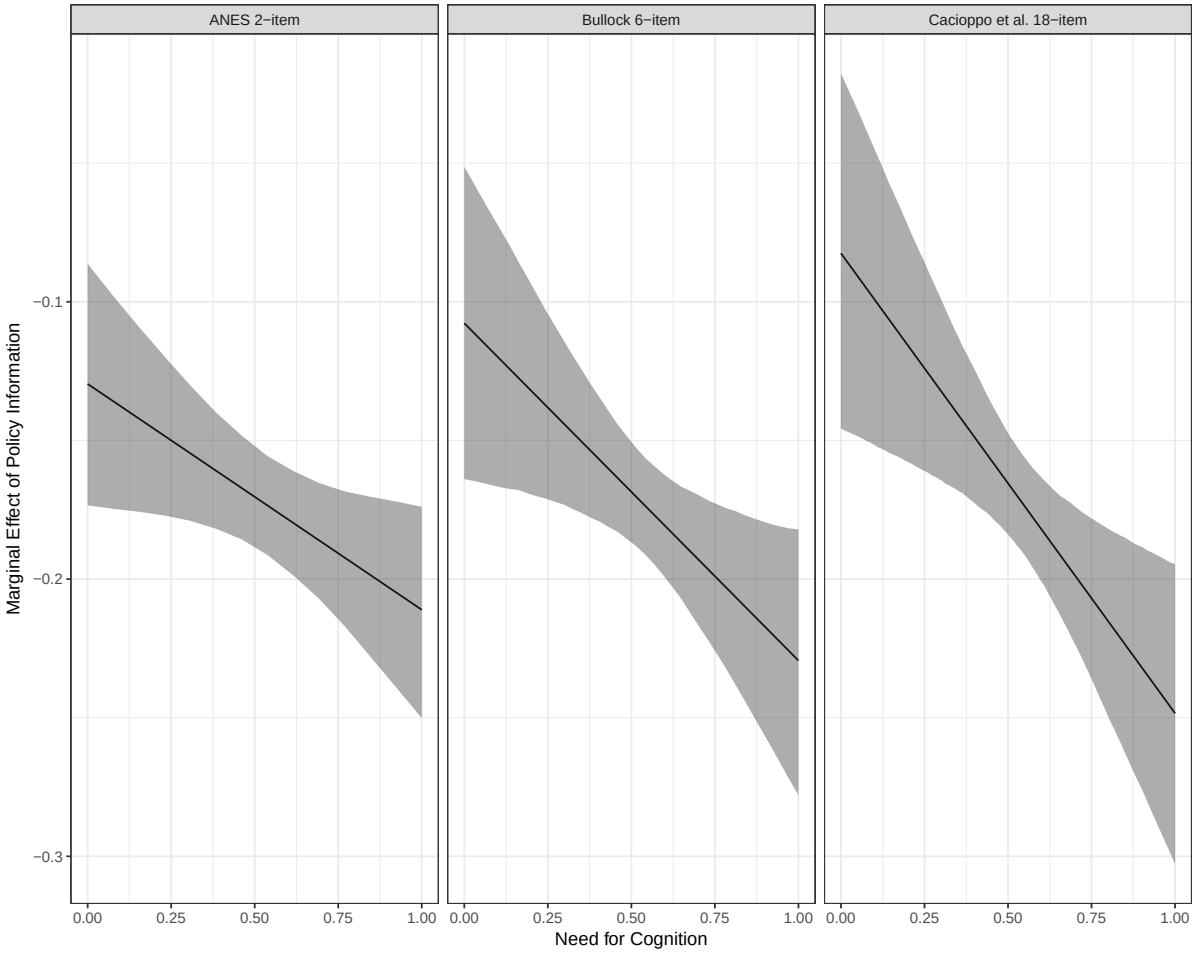
similar model using the 18-item NfC battery, a significant interaction effect that is more than twice the size of the ANES-based estimate of the interaction effect emerges (b=-0.17, SE=0.06). Specifically, the difference in the agreement between the right-wing and radical-right messages increases as NfC goes up (see right-hand panel of Figure 3). Subsetting our 18-item measure to the 6-item measure employed by Bullock (2011), yields a result that is slightly stronger than the 2-item measure (b=-0.12, SE=0.05), as can be seen in the middle panel of Figure 3. The results for the 6-item and 18-item NfC do not change once we directly account for measurement error (Appendix B.5). When we analyze the effect among liberals and conservatives separately, the NfC x message interaction is never significant among liberals, but among conservatives, the results are similar to those reported in Figure 3 (Appendix B.6). This is not surprising given that the message was either center-right (and appealing to conservatives) or far-right.

Next, we estimate whether the effect size is a function of the particular items or the number of items included in the battery (Figure 4). As in Study 1, these results clearly show the impact of scale length on the size of the effect. A one-item scale yields estimates of an interaction effect that is roughly a third of the 18 item scale. The median point estimate of randomly chosen scales does not increase linearly with the number of scales, however. The median point estimate of a randomly-chosen four item scale is three-fourths the size of the 18-item scale, by 9 items, the median coefficient estimate that is roughly 90 percent of the 18-item NfC estimate. The ANES measure seems to yield estimates that are smaller than roughly 70 percent of any other two-item measure indicating that the ANES measure is a particularly poor alternative to the full NfC measure.

As in study 1, we find that high Cronbach's alpha (Appendix B.7) or factor loadings (Appendix B.8) do not lead to measures of personality with a high criterion validity. Instead, Study 2 reaffirms that decreasing measurement error by increasing the number of items in a battery is the best strategy.

The NfC is a low bandwidth measure, and therefore differences in the criterion validity between the three NfC measures are most likely due to regression dilution, and

Figure 3. Need for Cognition and policy information



Note: Marginal effects of the radical-right message compared to the right-wing message with 95% Confidence Intervals are plotted (Appendix B.4 provides tables with results)

our errors were namely one of magnitude (i.e., Type M; Gelman & Carlin, 2014). Some personality traits – such as the Big Five traits – are much more complex and have a high bandwidth. What are consequences of using short measures for these constructs?

Short measures with high bandwidth: The Case of the Association between Personality and Political Ideology

Traditionally, the Big Five traits Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism are high bandwidth constructs that were assessed with up to 240 item scales using convenience samples of university students. However, more and more political scientists rely on short measures of the Big Five personality traits.

Figure 4. Need for Cognition and policy information

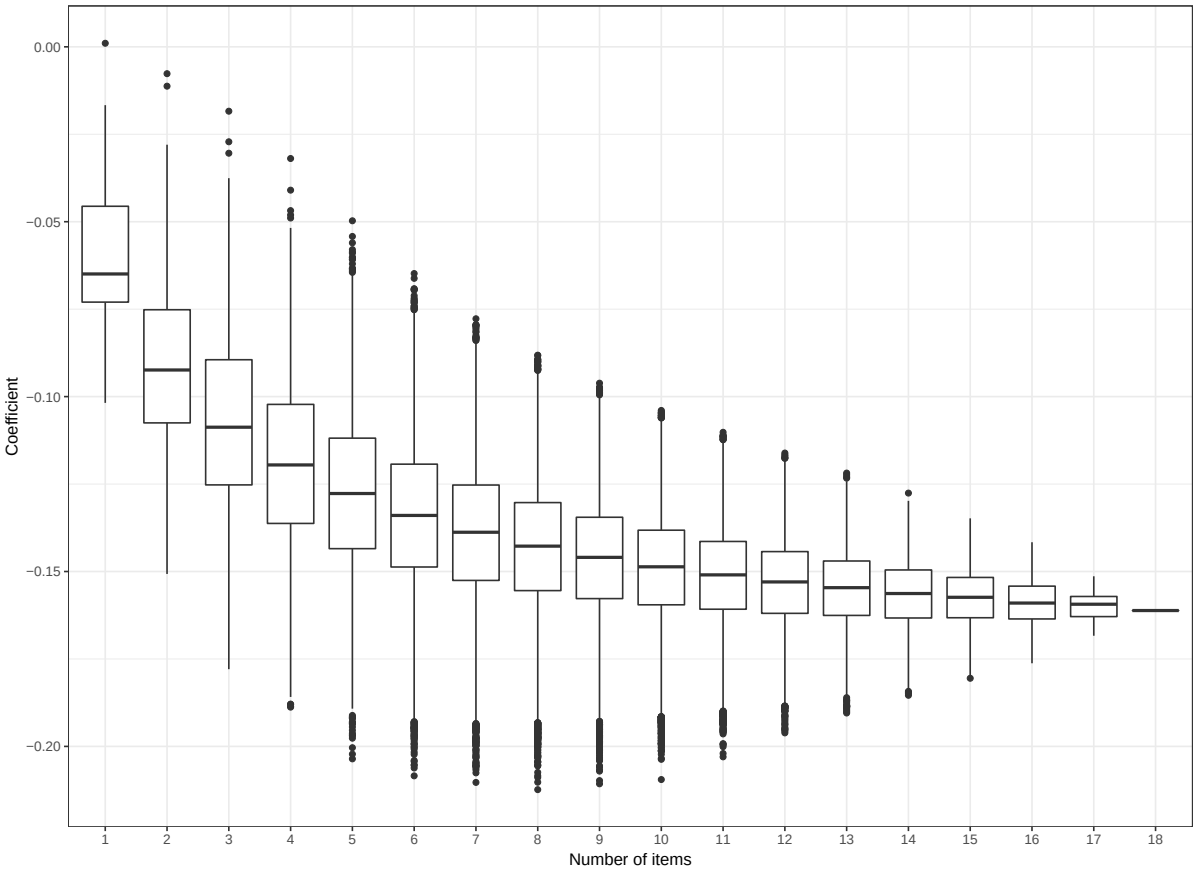


Figure plots the coefficient of the interaction effect between NfC and the policy information and each possible combinations of the NfC. Distribution of the point estimates of these measures sorted by the number of items used to generate the NfC.

Ten item personality inventories such as the BFI and TIPI are now included in the General Social Survey, the International Social Survey Programme, World Values Survey, the Cooperative Congressional Election Study, the American National Election Study, the AmericasBarometer and the British National Election Studies. The items for a short measure of each Big Five trait are selected so that the short measure reflects the breadth of the original dimension. Accordingly, the inter-item correlation is low (Gosling et al., 2003). Moreover, it is difficult to capture all aspects of a high bandwidth trait using only two items per trait. Necessarily, some aspects of a trait will be underrepresented in a short measure, which limits the content validity of the trait (Credé et al., 2012; Smith et al., 2000).

The burgeoning literature on the association between personality and political ideology (e.g., Gerber et al., 2010; Mondak & Halperin, 2008) could be particularly prone to the detrimental consequences of using brief measures. Gerber, Huber, Doherty, and Dowling (2011) indicated that correlations between political ideology and the traits Neuroticism, Agreeableness and Extraversion were consistently larger when measured with the TIPI compared to the 44-item Big Five Inventory (BFI), while the results for Openness and Conscientiousness varied to a lesser degree based upon the measure.[8] Although Gerber, Huber, Doherty, and Dowling (2011, p. 280) ultimately contend that differences between the TIPI and BFI were minor, they also conclude that "researchers should be sensitive to the consequences of using different personality batteries for predicting political outcomes."[9]

In order to generate a more complete comparison of Big Five measurements and outcomes, we surveyed the published literature (overview in Appendix C.1). These studies have yielded fairly consistent results when it comes to direction and statistical significance (although not strength) for negative association between Openness and conservatism as well as the positive association between Conscientiousness and conservatism (Appendix C.1, Table C1). Yet, our literature review shows that – with the exception of the consistent negative association between Openness and cultural conservatism (Appendix C.1, Table C2) – there is a heterogeneous pattern of associations between the Big Five traits and cultural conservatism whereby some studies find an association between the Big Five traits and others do not. Likewise, for economic ideology, more than half the studies for each trait point to an association with economic ideology, while between 30 and 50 percent of the studies failed to find an association (Appendix C.1, Table C3). With the exception of Openness, all traits have,

[8] Overestimation – just like underestimation – can occur when relying upon brief measures of personality (Credé et al., 2012).

[9] Mondak, Hibbing, Canache, Seligson, and Anderson (2010, Appendix A) partly address the issue by creating ten possible two item measures out of the 5-item measure of each trait. Yet, it is not possible to assess whether their results comport with underestimation or overestimation of the association because they group both underestimation and overestimation in one category.

to different degrees, been regarded as irrelevant for political ideology. Importantly, in the published literature, studies that relied upon brief measures – as well as convenience samples – are overrepresented in the studies that report no associations between traits and ideology.

Study 3: Big Five Traits and Political Ideology

Like in Study 2, we rely upon the Dutch LISS panel. In December 2012, a random subset of the LISS panelists (N=2,479) was invited to participate in the Dutch module of the 2012 World Values Survey (henceforth, WVS).[10] The response rate of the WVS was 76.6% (N=1,901) and the completion rate was 76.0% (N=1,884).

We combine the WVS with the "Personality and Values module" of the LISS panel which is fielded to LISS panel members in May of each year and contains a measure of the Big Five personality traits. To increase our sample size, we include the Big Five from May 2011 (76.4% response rate) and May 2012 (79.6% response rate).[11] Our sample was restricted to those participants who completed the WVS and had at least once completed a personality inventory in 2011 or 2012, this results in a data-set with 1,419 respondents.

Personality was assessed using two different batteries–the 10-item Big Five Inventory (BFI; Rammstedt & John, 2007) and the 50-item International Personality Item Pool – Five Factor Model (IPIP-FFM; Goldberg et al., 2006). When measuring Big Five traits there are roughly two traditions: inventories that consist of a series of questions on which respondents rate themselves and inventories that consist of a set of adjectives on which respondents rate themselves. The IPIP is part of the question based approach, while BFI is an adjective based approach. However, the IPIP-FFM and BFI show good convergent validity (Donnellan et al., 2006). A unique aspect of the 50-item IPIP-FFM is that it possible to derive a validated and reliable 20-item instrument, the Mini-IPIP (Donnellan et al., 2006).

The BFI and IPIP-FFM were assessed among panel members in different waves.

[10]Data collection period ran from December 3 2012 till December 31 2012.

[11]Data collection ran from May 2 till June 29 2011 and from May 7 and June 26 2012.

The BFI was administered as part of the WVS 2012. Participants completed the 50-item IPIP-FFM in the "Politics and Values" waves in May 2011 and May 2012. The items of both inventories were translated into Dutch by professional translators, using the translation-back-translation method, while the principle investigators of the panel resolved inconsistencies in the translations. The BFI was measured at the same time as the criterion measures, while the IPIP was measured *prior* to the criterion measures. Since personality traits are relatively stable over shorter time periods (Gerber et al., 2013) – and are stably associated with political attitudes over time (Bloeser, Canache, Mitchell, Mondak, & Poore, 2015) – the lag between the waves should not affect the nature of the associations reported here. Moreover, the strength of the associations between personality and the criterion measures should be biased in favor of the BFI compared to the measures collected earlier. We created additive scales of each of the Big Five traits.[12]

We compared the relationships between the personality traits, on the one hand, and different dimensions of political ideology, on the other, as these interrelationships are among the primary focus of the personality-politics research. Specifically, we focus upon a uni-dimensional operationalization of ideology (Mondak & Halperin, 2008) as well as cultural and economic ideology (Bakker, 2016; Feldman & Huddy, 2014; Gerber et al., 2010). Uni-dimensional ideology was part of the WVS 2012 and measured by asking panelists to rate themselves on a scale from left (0) to right (10). We recoded the ideology dimension to range from the most liberal (0; left) to most conservative (1; right) observation (M = 0.51, SD = 0.23).

Cultural and economic ideology were measured using a total of nine items that were part of the WVS 2012.[13] Cultural ideology was measured using six items: "I find it shocking if two men kiss in public"; "Gay men and lesbian women should be free to live their life as they wish" and "I find it shocking when a man and a woman kiss in

[12]See Appendix C.2 for item wording and Appendix C.3 for the psychometric characteristics of the different personality batteries.

[13]Three items were included from the "Politics and Values" wave which was completed by LISS panelists at the same time as the WVS (i.e., December 2012).

public" all scored on a five point Likert-scale ranging from "strongly disagree" (1) through "strongly agree" (5). Abortion "can always be justified" (1) through "never be justified" (10); "Where would you place yourself on a scale from 1 to 5, where 1 means that euthanasia should be forbidden and 5 means that euthanasia should be permitted"; "There are too many people of foreign descent in the Netherlands." Scored on a scale ranging from "fully disagree" (1) to "fully agree" (5).

Economic ideology was measured using three items: "Incomes should be made more equal" (1) through "Individual effort should be rewarded" (10); "Government should take more responsibility to ensure that everyone is provided for" (1) through "People should take more responsibility to provide for themselves" (10); "Where would you place yourself on a scale from 1 to 5, where 1 means that differences in income should increase and 5 means that these should decrease?"

A confirmatory factor analysis confirmed that the nine items fit within a two-dimensional factor structure whereby the items load high on the hypothesized ideology dimension (Appendix C.4). The internal consistency of the cultural ideology dimension ($\alpha = 0.70$) and economic ideology ($\alpha = 0.75$) were acceptable. We created a scale ranging from the most liberal (0) through the most conservative (1) cultural ideology (M = 0.33; SD = 0.16) and a scale ranging from the most liberal (0) through the most conservative (1) economic ideology (M = 0.45, SD = 0.20).

The three ideology dimensions are conceptually distinct. The correlation between unidimensional ideology and cultural ideology was fairly weak (r=0.29), while the correlation between unidimensional ideology and economic ideology was modestly strong (r=0.54). Cultural and economic ideology were weakly associated with each other (r=0.10) (see also, Bakker, 2016; Feldman & Johnston, 2014).

For each personality battery, we regressed – using OLS regression models – the criterion measures on each trait as well as sex, age, education and income (see Appendix C.5 for the descriptive statistics). To make the results easily comparable, we plot the unstandardized regression coefficients and 95 percent confidence intervals in one figure with a row for each trait and a column for the associations between each

ideology dimension and the BFI, Mini-IPIP and IPIP. We discuss the results for the three ideology measures on a trait by trait basis.[14] The results discussed here do not change if we do not control for education and income (Appendix C.8) or if we run structural equation models (Appendix C.9).

Starting with Openness, higher levels of Openness were negatively correlated with conservatism (column 1) and cultural conservatism (column 2) (Figure 5). However, a study utilizing the BFI would conclude that there is a small negative association between economic ideology and Openness, while a study utilizing the IPIP would conclude that there is no association between economic ideology and Openness (Figure 5, row 1, column 3).

Turning to Conscientiousness (Figure 5, row 2), there is a consistent positive association between Conscientiousness and the unidimensional measure of conservatism (row 2, column 1) cultural conservatism (row 2, column 2). However, the association between Conscientiousness and these ideology dimensions was 1.5-2 times larger when measured using the IPIP compared to the BFI which suggests an underestimation of the effect of Conscientiousness. Finally, measurement of Conscientiousness conditions the conclusion we reach about the relationship between Conscientiousness and economic conservatism (Figure 5, row 2, column 2 & 3). We would likely conclude that there was no relationship between Conscientiousness and the economic ideology dimensions because the 95 percent confidence intervals of the BFI contains zero. Using the more elaborate IPIP and Mini-IPIP, Conscientiousness and economic conservatism are positive and significantly correlated. Thus scholars using the BFI underestimate the size of the association between Conscientiousness and ideology and might even conclude that there is no association between Conscientiousness and economic ideology.

The results for Neuroticism show that the trait is consistently correlated with the ideology dimensions (Figure 5, row 3). There is no association between Neuroticism and a uni-dimensional measure of ideology, while Neuroticism is positively correlated with cultural conservatism and negatively correlated with economic conservatism. Our

[14] Standardized regression coefficients are reported in Appendix C.7.

results suggest that the measurement of Neuroticism does not condition the conclusions we reach about the association between this trait and various ideology dimensions.

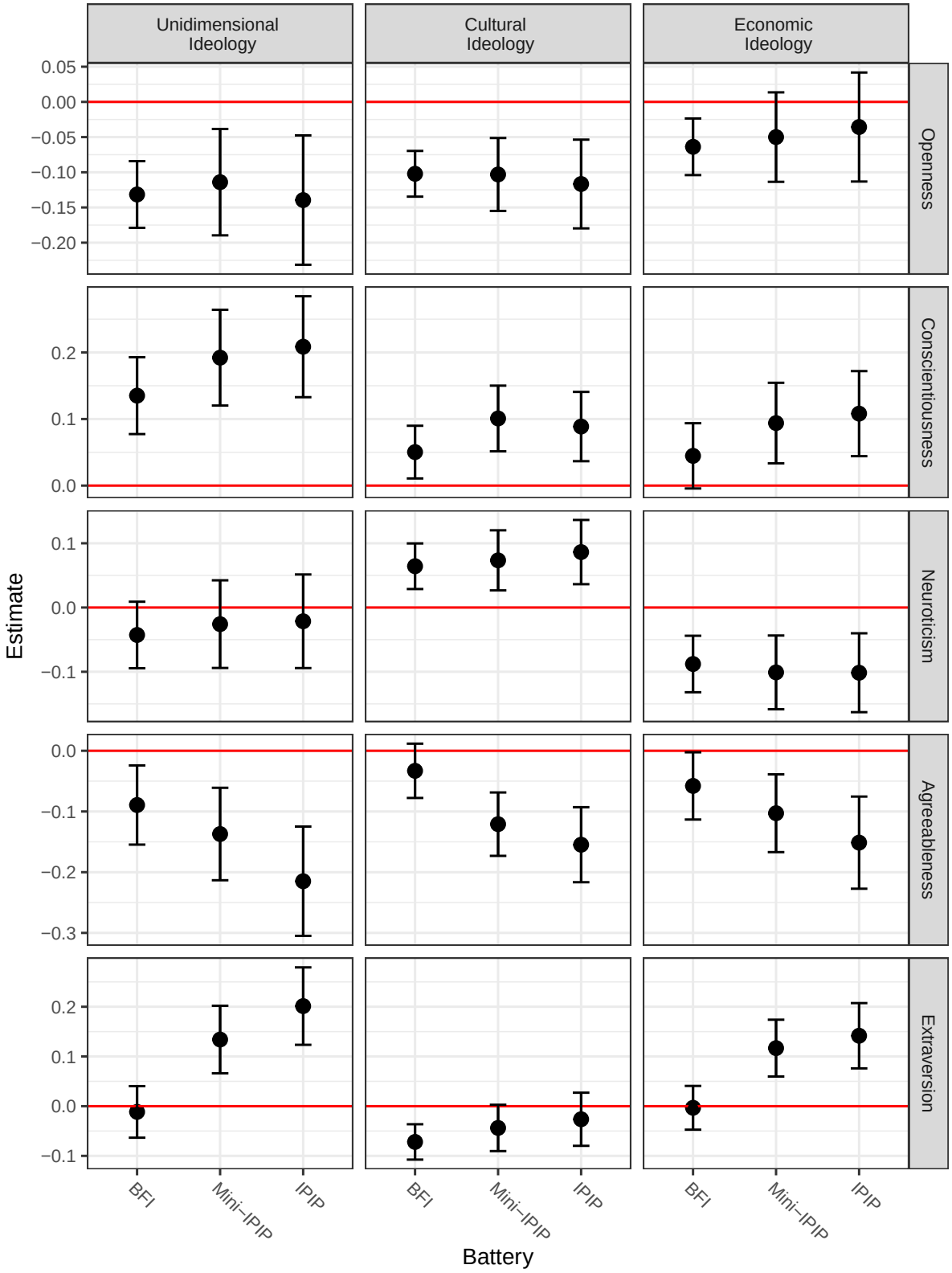
Measurement conditions the substantive conclusions we draw about the association between Agreeableness and the dimensions of ideology (Figure 5, row 4). The negative association between Agreeableness and conservatism (row 4, column 1) is two times as large using the IPIP compared to the BFI, while the Mini-IPIP estimate is roughly 50% larger compared to the BFI although not statistical significant different.

The BFI estimate of the relationship between Agreeableness and cultural conservatism was not significant. Yet, this seems to be a Type M error, the relationship between Agreeableness and cultural conservatism is *three* times larger compared to the IPIP. Finally, if we employ the BFI in the study of economic ideology, we would likely conclude that there is a weak negative association with economic conservatism. The negative association between Agreeableness and economic conservatism is roughly three times as large and significant when we use the IPIP. The Agreeableness results indicate that there is a severe risk of a Type M error when using a brief measure of the trait.

Extraversion yielded striking results with both Type M and Type S errors occurring. Type M errors are found for the relationship between Extraversion and an unidimensional measure of ideology as well as cultural ideology. Using the BFI, the association between Extraversion and unidimensional ideology was negative and not different from zero. The IPIP - as well as the Mini-IPIP - estimate was positive and much larger and statistical significantly stronger compared to the BFI estimate. Similarly, those using the BFI would conclude that the relationship between cultural ideology and Extraversion was negative and not significant, those using the IPIP would probably argue that there is no relationship between the two constructs. Finally, a Type S error is found for the relationship between Extraversion and cultural ideology. Those using the BFI would find a negative but not different from zero relationship between Extraversion and economic ideology. Those using either the Mini-IPIP or the IPIP would find a much larger significantly positive one.

Perhaps another brief personality inventory would results in estimates more

Figure 5. Personality and politics: BFI, Mini-IPIP and IPIP-FFM results



Note: unstandardized OLS estimates with 95% Confidence Intervals are plotted (see Appendix C.6 for tables with results).

consistent with larger personality batteries? Unfortunately, our study does not contains alternative brief measures. We can, however, following the same logic employed in study 1 and 2, generate 10 1-item measures, 45 different 2-item measures, 120 3-item measures, 210 4-item measures, 252 5-item measures, 210 6-item measures, 120 7-item measures, 45 8-item measures, and 10 9-item measures. We calculated the associations between our measures of ideology and each of these 1,022 possible combinations of the trait, controlling for the 10-item measures of the other four traits.[15] Figure 6 plots the distribution of the point estimates of these measures sorted by the number of items used to generate the trait. These results clearly illustrate that decreasing the number of items–regardless of the items chosen–generally attenuates the relationship between a trait and the ideology dimensions.

Finally, and in line with study 1 and 2, we show that selecting items of a scale using Cronbach's Alpha (Appendix C.10) or factor loadings (Appendix C.11) does not lead to better estimates.

Discussion

The current "replication crisis" has raised serious questions about the validity of social science findings. Political scientists may feel shielded because sample sizes tend to be much larger and more representative than those in psychology, particularly when we use large-N surveys like the American National Election Studies, Cooperative Congressional Election Studies or the World Values Survey. However, as a trade-off for large sample sizes, political scientists rely on very short measures of various constructs. As we have demonstrated, this trade-off often leads to different conclusions than if our constructs of interest were longer (see also, Achen, 1975; Ansolabehere et al., 2008).

The general consensus in political science is that NfC does not moderate a person's reliance on cues – a consensus that runs counter to the Elaboration Likelihood

[15] We control for the covariates as well as 10-item measure of the other four traits because this yields the most conservative test given that we reduce measurement error. It becomes logistically difficult to estimate all possible covariate combinations with all possible criterion combinations as it produces over 1.1×10^{15} parameters.

Figure 6. Study 3: Relationship between personality and political ideology based upon the number of items used to create each Big Five trait

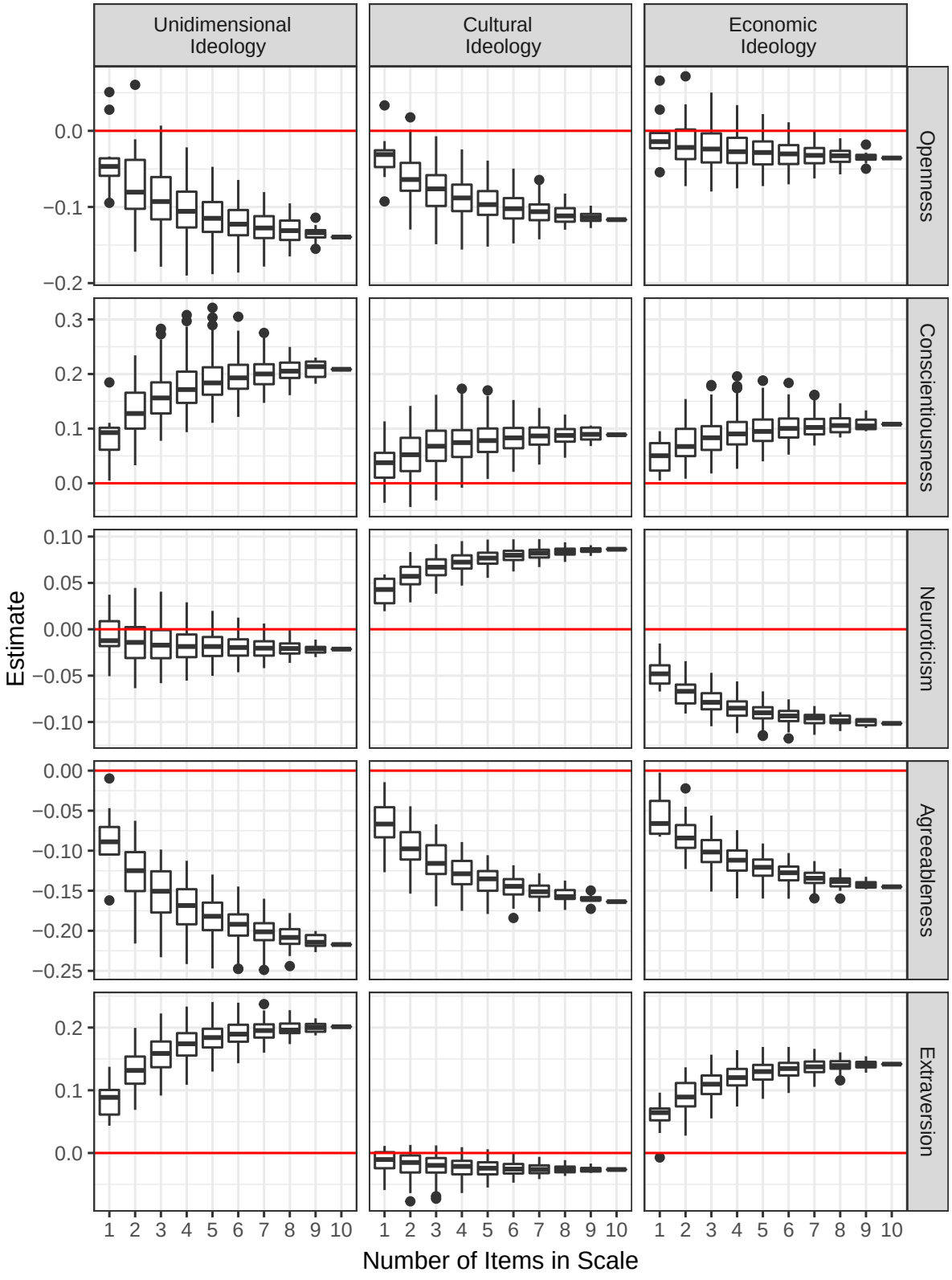


Figure plots the associations between our measures of ideology and each possible combination of each trait. Distribution of the point estimates of these measures sorted by the number of items used to generate the trait.

Model (Petty & Cacioppo, 1986). This work overwhelmingly relies on the 2-item ANES measure, but when longer measures are used, we have shown in study 2 that those with higher levels of NfC are more likely to rely upon policy information. Additionally, our first study indicates that those who exhibit higher levels of NfC are also more likely to rely on party cues, a finding that runs counter to the Elaboration Likelihood Model and Kam's (2005) expectations, but which is consistent with theories of motivated reasoning (Kahan, 2012; Slothuus & De Vreese, 2010).

Turning to the Big Five and ideology literature, we have shown that – with the exception of Neuroticism – the association between personality and political dimensions is highly conditional upon the measurement of personality. We found that 50-item IPIP-FFM yields associations with ideology that are twice as strong as the associations produced by the BFI. In a few instances, the BFI yields estimates of the opposite sign than the 50 item measure which suggests the possibility of a Type S error. Traits which have largely been dismissed as irrelevant for the study of politics and personality – such as Extraversion and Agreeableness – are as strongly correlated with our outcome measures as those which are focal to the field. Our study thus shows that relying upon a larger Big Five battery would yield different conclusions about which traits are correlated with political ideology.

This study is not without its limitations. One could argue that differences in criterion validity between the BFI and the IPIP in study 3 are actually not explained by the length of the battery but by the measurement tradition, i.e., adjectives versus sentences. However, our randomly generated two item IPIP measures in Figure 6 show the same poor criterion validity as the adjective based BFI. Yet, in order to rule out this alternative explanation completely future studies should collect data that contains brief and elaborate measures that are based upon the questionnaire and adjective approach.

Additionally, when asking questions about the criterion validity, studies ideally utilize some gold standard, wherein they compare some self-reported behavior with an actual behavior such as the study of electoral participation (for instance, Gerber, Huber, Doherty, Dowling, Raso, & Ha, 2011). We do not have an analogous criterion

measure here. This implies that we have to be careful in drawing conclusions that the results of the larger batteries results in *better* estimates. We have reasons to believe that the results of the larger batteries leads to estimates closer to the true estimate because of the superior measurement properties. Yet, we have no way of proving this point. More research, using independent samples, but equivalent measures should help us to get one step closer to understanding the size and direction of the association between personality and political ideology. Finally, many studies of personality and politics have been conducted among American subjects, and it is possible that our results may have been influenced by cultural differences because study 2 and 3 were conducted in the Netherlands. Future research is well advised to replicate and extend these findings in other political contexts such as the United States.

The renewed interest in personality has also sparked interest in a host of other personality inventories that we did not investigate such as other popular Big Five inventories (i.e., the TIPI, Gosling et al., 2003) or other popular personality inventories such as the Need to Evaluate and the Need for Certainty. There is no a priori reason to assume that other brief measures of personality are less prone to the Type M – and to some extent Type S – errors documented in this study. Yet, the conclusions in this study are necessarily limited to the brief measures of personality we employed. We would welcome future research that assesses the consequences of the use of these brief personality measures because it is not possible to generalize our findings to other brief measures of personality without a direct empirical test.

Finally, identifying the particular combination of items which balance validity and scale length is beyond the scope of this paper. But we can offer a few suggestions for future research. As discussed by Cronbach and Meehl (1955, p. 300), "many types of evidence are relevant to construct validity, including content validity, inter-item correlations, inter-test correlations, test-criterion correlations, studies of stability over time, and stability under experimental intervention." Scholars tend to follow Cronbach and Meehl (1955) and assess the inter-item correlations (see for instance, Gerber et al., 2010) and over-time stability of the construct (Gerber et al., 2013). However, more

attention should be paid to the criterion validity of a construct (for notable exceptions see Gerber, Huber, Doherty, & Dowling, 2011; Kam & Estes, 2016).

What is the way forward for the personality and politics literature in political science? Adaptive tests of personality (Montgomery & Cutler, 2013) – such as the implementation of the NfC in the 2016 ANES pilot – save considerable space in surveys. However, some respondents will still answer long measures which is costly, time-consuming and tiring. Another possibility is to randomly assign k items from the larger pool of N items. Given that this fulfills the missing completely at random assumption, we can then impute missing values so that a scale of length N is generated for each respondent (Little & Rhemtulla, 2013).

Finally, we believe there is still value in utilizing the abbreviated measures that have been implemented on omnibus surveys. The longitudinal and cross-national nature of many of these surveys make them particularly invaluable. As evidenced by the NfC studies, short measures of low-bandwidth traits seem to accurately gauge the direction of the effect if not the magnitude. Hence, we should be more skeptical of null relationships from analyses that use these types of measures than non-null relationships from that use these measures. One way forward would be to replicate results found using high quality, probability samples and brief measures on (less expensive) non-probability samples samples with longer measures, thereby ensuring both external validity and internal validity.

Moving forward, we strongly advise against employing the 2-item ANES NfC measure to study the reliance upon party cues or policy information. Regression dilution seems to seriously affect our results and the conclusions we draw. However, the 18-item NfC measure seems to be overkill. A randomly drawn 6-8 item battery performs more or less in line with the 18-item measures. While we have to be careful to generalize based upon this study, we strongly advise researchers to devote some more space in their survey for a NfC battery (see for instance, Arceneaux & Vander Wielen, 2017; Bullock, 2011). Our results suggests that conventions such as high factor loadings and/or high Cronbach's alpha should not per se be the yardsticks to select the items. Instead,

randomly generating a battery of 6-8 items would suffice. Yet, we do urge scholars to pretest their NfC battery paying close attention to the criterion validity of their scale.

Like the NfC, we advise against the use of brief measures of the Big Five personality traits such as the 10-item Big Five Inventory. Type M and even Type S errors seem to dominate these findings. Accordingly, scholars run the risk to disregard traits as relevant to politics, while they are in fact relevant. Again, a randomly drawn 6-8 item battery per trait – or to some extent relying upon the 4-item Mini-IPIP – seems to function in line with the larger 10-item battery.

We believe the literature on personality and politics has arrived at a turning point and should move beyond the use of extremely abbreviated scales of personality. We therefore welcome the next of generation of personality and politics research.

Acknowledgements

Both authors contributed equally to this paper. Earlier versions of this manuscript were presented at Tilburg University, the University of Amsterdam, the Midwest Political Science Association Meeting (2015) and the International Society of Political Psychology Meeting (2016). We thank Mark Brandt, John Bullock, Scott Clifford, Donald Green, Cindy Kam, Gerard Saucier, Martijn Schoonvelde, Kevin Smith, Marco Steenbergen, Claes de Vreese, Mariken van der Velden, the five anonymous reviewers and editor Jennifer Merolla for their comments and suggestions that helped us improve the paper.

References

Achen, C. H. (1975). Mass political attitudes and the survey response. *American Political Science Review*, *69*(4), 1218-1231.

Alvarez, R. M., & Brehm, J. (1995). American ambivalence towards abortion policy: Development of a heteroskedastic probit model of competing values. *American Journal of Political Science*, *30*(4), 1055-1082.

Ansolabehere, S., Rodden, J., & Snyder, J. M. (2008). The strength of issues: Using multiple measures to gauge preference stability, ideological constraint, and issue voting. *American Political Science Review*, *102*(2), 215-232.

Arceneaux, K., & Vander Wielen, R. J. (2017). *Taming intuition: How reflection minimizes partisan reasoning and promotes democratic accountability.* Cambridge University Press.

Bakker, B. N. (2016). Personality traits, income and economic ideology. *Political Psychology*. doi: 10.1111/pops.12349

Bakker, B. N., Rooduijn, M., & Schumacher, G. (2016). The psychological roots of populist voting: Evidence from the united states, the netherlands and germany. *European Journal of Political Research*, *55*(2), 302–320.

Berinsky, A. J., & Kinder, D. R. (2006). Making sense of issues through media frames: Understanding the kosovo crisis. *Journal of Politics*, *68*(3), 640-656.

Bizer, G. Y., Krosnick, J. A., Petty, R. E., Rucker, D. D., & Wheeler, S. C. (2000). Need for cognition and need to evaluate in the 1998 national election survey pilot study. *National Election Studies Report*.

Bloeser, A. J., Canache, D., Mitchell, D.-G., Mondak, J. J., & Poore, E. R. (2015). The temporal consistency of personality effects: Evidence from the british household panel survey. *Political Psychology*, *36*(3), 331-340.

Bullock, J. G. (2011). Elite influence on public opinion in an informed electorate. *American Political Science Review*, *105*(3), 496-515.

Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. *American Psychologist*, *39*(3), 214-227.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116-131.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306-307.

Coffe, H. (2013). It's all about the party: Gender, party characteristics, and radical right voting.

Costa, P. T., McCrae, R. R., & Dye, D. A. (1991). Facets scales for agreeableness and conscientiousness: A revision of the NEO personality inventory. *Personality and Individual Differences*, *12*(9), 887-898.

Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the big five personality traits. *Journal of Personality and Social Psychology*, *102*(4), 874-888.

Cronbach, L. J. (1949). Essentials of psychological testing.

Cronbach, L. J., & Gleser, G. C. (1965). Psychological tests and personnel decisions.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality. *Psychological Assessment*, *18*(2), 192-203.

Edwards, P., Roberts, I., Sandercock, P., & Frost, C. (2004). Follow-up by mail in clinical trials: does questionnaire length matter? *Controlled Clinical Trials*, *25*(1), 31-52.

Feldman, S., & Huddy, L. (2014). Not so simple: The multidimensional nature and diverse origins of political ideology. *Behavioral and Brain Sciences*, *37*(3), 312–313.

Feldman, S., & Johnston, C. (2014). Understanding the determinants of political ideology: Implications of structural complexity. *Political Psychology*, *35*(3), 337-358.

Gelman, A., & Carlin, J. (2014). Beyond power calculations assessing type s (sign) and

type m (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641-651.

Gerber, A. S., Huber, G. A., Doherty, D., & Dowling, C. M. (2011). The big five personality traits in the political arena. *Annual Review of Political Science*, *14*(1), 265–287.

Gerber, A. S., Huber, G. A., Doherty, D., & Dowling, C. M. (2013). Assessing the stability of psychological and political survey measures. *American Politics Research*, *41*(1), 54–75.

Gerber, A. S., Huber, G. A., Doherty, D., Dowling, C. M., & Ha, S. E. (2010). Personality and political attitudes: Relationships across issue domains and political contexts. *American Political Science Review*, *104*(1), 111-133.

Gerber, A. S., Huber, G. A., Doherty, D., Dowling, C. M., Raso, C., & Ha, S. E. (2011). Personality traits and participation in political processes. *The Journal of Politics*, *73*(3), 692-706.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, *37*(6), 504–528.

Hatemi, P. K., & McDermott, R. (2016). Give me attitudes. *Annual Review of Political Science*, *19*, 331-350.

Holbrook, T. M. (2006). Cognitive style and political learning in the 2000 u.s. presidential campaign. *Political Research Quarterly*, *59*(3), 343-352.

Johnson, T. R., & Martin, A. D. (1998). The public's conditional response to supreme court decisions. *American Political Science Review*, *92*(2), 299-309.

Kahan, D. M. (2012). Ideology, motivated reasoning, and cognitive reflection: an experimental study. *Judgment and Decision Making*, *8*, 407-424.

Kam, C. D. (2005). Who toes the party line? cues, values, and individual differences. *Political Behavior*, *27*(2), 163-182.

Kam, C. D., & Estes, B. A. (2016). Disgust sensitivity and public demand for protection. *The Journal of Politics*, *78*(2), 481-496.

Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, *7*(4), 199–204.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Welsley Publishing Company.

Malka, A., Soto, C. J., Inzlicht, M., & Lelkes, Y. (2014). Do needs for security and certainty predict cultural and economic conservatism? a cross-national analysis. *Journal of personality and social psychology*, *106*(6), 1031–1051.

Mérola, V., & Hitt, M. P. (2016). Numeracy and the persuasive effect of policy information and party cues. *Public Opinion Quarterly*, *80*(2), 554-562.

Mondak, J. J., & Halperin, K. D. (2008). A framework for the study of personality and political behaviour. *British Journal of Political Science*, *38*(2), 335-362.

Mondak, J. J., Hibbing, M. V., Canache, D., Seligson, M. A., & Anderson, M. R. (2010). Personality and civic engagement: An integrative framework for the study of trait effects on political behavior. *American Political Science Review*, *104*(1), 85-110.

Montgomery, J. M., & Cutler, J. (2013). Computerized adaptive testing for public opinion surveys. *Political Analysis*, *21*(2), 172-192.

Petersen, M. B., Skov, M., Serritzlew, S., & Ramsøy, T. (2013). Motivated reasoning and political parties: Evidence for increased processing in the face of party cues. *Political Behavior*, *35*(4), 831-854.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Communication and persuasion* (p. 1-24). Springer.

Popkin, S. L. (1994). *The reasoning voter: Communication and persuasion in presidential campaigns*. University of Chicago Press.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of Research in Personality*, *41*(1), 203-212.

Rudolph, T. J. (2011). The dynamics of ambivalence. *American Journal of Political*

Science, *55*(3), 561-573.

Scherpenzeel, A., & Das, J. (2010). "true" longitudinal and probability-based internet panels: Evidence from the netherlands. *Social and behavioral research and the internet: Advances in applied methods and research strategies*, 77-103.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment*, *8*(4), 350–353.

Slothuus, R., & De Vreese, C. H. (2010). Political parties, motivated reasoning, and issue framing effects. *The Journal of Politics*, *72*(3), 630-645.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*(1), 102-111.

Sokhey, A. E., & McClurg, S. D. (2012). Social networks and correct voting. *The Journal of Politics*, *74*(3), 751-764.

Tidwell, P. S., Sadowski, C. J., & Pate, L. M. (2000). Relationships between need for cognition, knowledge, and verbal ability. *The Journal of Psychology*, *134*(6), 634-644.

Bibliography

Bert N. Bakker is an assistant professor at the University of Amsterdam, Amsterdam (The Netherlands), 1001 NG and Visiting Research Fellow at Temple University, Philadelphia (United States) PA 19122.

Yphtach Lelkes is an assistant professor at the University of Pennsylvania, Philadelphia (United States), PA 19104.